




Specific Emitter Identification Based on ACO-XGBoost Feature Selection

Jianjun Cao¹, Chumei Gu^{1,2}(✉) , Baowei Wang², Yuxin Xu^{1,2}, and Mengda Wang^{1,2}

¹ The Sixty-Third Research Institute, National University of Defense Technology, Nanjing, China

caojj@nudt.edu.cn, m15261820030@163.com

² School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China

Abstract. In order to improve the performance of Specific Emitter Identification (SEI) in complex electromagnetic environment, a SEI model based on Ant Colony Optimization-eXtreme Gradient Boosting (ACO-XGBoost) feature selection is proposed. Firstly, lifting wavelet package transformation is introduced to extract features of original emitter signals and data normalization is carried out on the data. Secondly, ACO is used to optimize the main parameters for XGBoost and the optimized XGBoost algorithm is applied to score the importance of the features. Thirdly, the importance of each feature is used as threshold to determine the best feature subset. Finally, combined with XGBoost classifier, specific emitter has been correctly identified. The proposed model is investigated on three datasets and the experimental result shows that the optimal feature subset can be effectively acquired and reasonably classified by the proposed model. Comparing with Decision Tree (DT), Random Forest (RF), Gradient Boosting Decision Tree (GBDT) and XGBoost feature selection methods, the identification performance of the proposed model is effectively improved.

Keywords: Lifting wavelet package transformation · Specific emitter identification · Ant colony optimization · XGBoost · Feature selection

1 Introduction

With the rapid development of communication engineering and the population of mobile communication device, wireless communication has more and more indispensable of modern society. The obtained signal from specific emitter not only carries signal information, but also contains the hardware information of the emitter, which is usually called fingerprint feature of emitter [1]. SEI method identifies different emitters based on fingerprint features which is detectable, immutable and tamper-resistant. SEI has great application value in military and civil fields. In the military field, it can fundamentally improve the investigation and confrontation ability against enemy emitter by extracting subtle feature of signal and identifying the emitter. In the civil field, this method can locate the emitter lacking detection and eliminate the fault timely. SEI is mainly divided

into three steps: signal preprocessing, feature extraction and classification [2]. However, the feature vector of emitter after using some preprocessing and feature extraction methods is high dimensional and low sample size. Data of these characteristics pose a challenge to effective classification and further analysis. One feasible and powerful method to solve the above problems is feature selection.

Feature selection (FS) assesses the importance score of every feature and constructs an optimal feature subset efficient for pattern recognition tasks [3]. The purpose of feature selection is to obtain the best predictive accuracy by selecting relevant features and removing irrelevant or redundant features [4]. Feature selection techniques usually be divided into three categories: filter, wrapper and embedded methods [5]. The filter methods evaluate the importance of data according to its inherent characteristics. Filter methods first use an evaluation function to score every feature and then all features are sorted. Finally, some top- k features are selected to constitute a feature subset. This method is simple, fast and independent of learning algorithm but ignores the dependence between features. Typical filter feature selection methods are Relief, Pearson Correlation Coefficient and Mutual Information (MI), etc. [6]. Wrapper methods wrap feature selection around learning algorithm, and achieve the feature subset with the most discriminative ability by minimizing the prediction error of classifier. This method considers the correlation between features, but indeed increases computation complexity and overfitting risk. Recursive Feature Elimination (RFE) is typical wrapper feature selection method [7]. Embedded methods embed feature selection into specified learning algorithm like Linear SVM, DT, RF, GBDT and XGBoost, etc. [8]. After training, the importance of each feature is achieved and then the optimal feature subset can be obtained. This method avoids the repeated execution of classifier.

Because embedded methods can be regarded as a balance between filter and wrapper methods, scholars have done lots of researches on feature selection based on embedded methods. Chen et al. [9] proposed a network intrusion detection model based on RF and XGBoost to improve the accuracy and real-time performance in complex network environment. The feature importance was calculated based on RF and a hybrid feature selection method combining filtering and embedded was used for feature selection. Then XGBoost method was used to detect and recognize the optimal feature subset. The model greatly reduced processing time with high detection accuracy. Zhou et al. [10] proposed a model to solve the multi-classification problem of unbalanced data in network intrusion detection. GBDT was used to calculate the importance of features and rank them, and Recursive Feature Elimination was used for feature selection. Then RF was used for feature conversion. The model had significant advantages in solving the multi-classification problem of unbalanced data. Based on the negative impact of imbalanced data on multi-classification problem, Feng et al. [11] proposed a stacked model based on XGBoost and RF feature selection. The new feature importance was the harmonic mean of the feature importance of RF and XGBoost. The threshold value was set of 0.01 and features less importance than this threshold is eliminate. This method effectively improved the classification performance. Gong et al. [12] proposed an interpretable Traditional Chinese Medicine treatment model based on XGBoost to overcome the problem of ignoring the trust mechanism of decision-making process. XGBoost

was used to calculate feature importance and filter non representative features according to feature importance. Besides, the model training combined base feature selection, parameter optimization and model integration. This model was considered interpretable in the feature selection and classification. To solve the problem of low neuropsychiatric disorder classification accuracy caused by single data type, Liu et al. [13] proposed an ensemble hybrid feature selection method. 3D DenseNet was used to select image features from magnetic resonance imagines and XGBoost was used to select phenotypic features from feature importance and then image features and phenotypic features were spliced in the form of vectors. The hybrid features improved the performance of classification algorithms. Qiao et al. [14] proposed an intrusion detection model based on the combination of XGBoost and RF to address the problem of inaccurate classification. XGBoost was used to score the importance of features and improved RF was use to judge whether the network traffic is normal or abnormal. The optimal features could be effectively selected and classified by this model.

According to the advantages and disadvantages of the above ideas, we propose a new method for SEI based on ACO-XGBoost feature selection. This method mainly combines lifting wavelet package transformation, ACO and XGBoost. The main contributions of this paper are summarized as follows:

- We use lifting wavelet package transformation to extract features. Then the feature parameter system is established.
- ACO algorithm is used to optimize the parameters of XGBoost. The paper relates the maximum depth of tree (χ), the minimum sum of instance weight needed in a child (δ), the number of decision trees (ε), the L2 regularization term on weights (γ), the minimum loss reduction required to make a further partition on a leaf node (λ) and learning rate (ω). The best parameters will be beneficial for SEI.
- The optimized XGBoost algorithm is used to obtain the importance of every feature and every importance is used as threshold to obtain the final best feature subset. XGBoost will also serve as a classifier to determine the classification result.
- Experiments on radio datasets in different states show the efficiency and effectiveness of the proposed model by comparing the performance of different feature selection methods, including DT, RF, GBDT and XGBoost.

The remainder of this paper is organized as follows. Section 2 introduces the model of SEI based on ACO-XGBoost FS. In Sect. 3, three specific emitter datasets are employed to prove the effectiveness and efficiency of the proposed method. Finally, we summarize this paper and propose the future work in Sect. 4.

2 Mathematical Model of Parameter Optimization and FS of XGBoost

In order to identify different specific emitters, the features should be extracted of collected original signals. The extracted features can be used to construct a feature set $set = \{v_i | v_i = v_1, v_2, \dots, v_n\}$, $i = 1, 2, \dots, n$. The feature vector corresponding to the features in the set is recorded as V , and the existing Y feature vector sample of class

W is recorded as V_{wyi} , $w = 1, 2, \dots, W$, $y = 1, 2, \dots, Y$, $i = 1, 2, \dots, n$, V_{wyi} is the i th feature value of the y th sample vector in class w . Some of the extracted features are highly relevant to the prediction result, but there are irrelevant and redundant features. The input of these features will greatly reduce the identification performance of specific emitters. Effective feature selection can reduce the feature dimension and improve the performance of SEI.

The essence of SEI is classification and embedded feature selection methods embed feature selection into specified learning algorithm. The selection of classifier parameters and feature subsets directly affect the final classification performance. In this paper, the classification accuracy of classifier and the number of features in the selected subset are directly used as the objective function. The mathematical model of parameter optimization and FS of XGBoost can be described as: selecting the value of six parameters of XGBoost and a subset $subset^q$ with cardinality q from the original feature set, where the classification accuracy A is maximal and the number of features q is minimal. The specific mathematical model are as follows:

$$\max A(\chi; \delta; \varepsilon; \gamma; \lambda; \omega; subset^q) \quad (1)$$

$$\min q \quad (2)$$

$$\text{s.t. } |subset^q| = q, 1 \leq q \leq n \quad (3)$$

where A is the classification accuracy and its equation is as follows:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

TP : the real category of the sample is positive, and the result of model recognition is also positive. FN : the real category of the sample is positive, but the result of model recognition is negative. FP : the real category of the sample is negative, but the result of model recognition is positive. TN : the real category of the sample is negative, and the result of model recognition is also negative. Accuracy represents the ratio of the number of correctly classified samples to the total number of samples.

3 SEI Based on ACO-XGBoost FS

The basic framework of SEI based on ACO-XGBoost FS is shown in Fig. 1. Firstly, the features are extracted based on lifting wavelet package transformation, then the original feature data set is formed. Secondly, the feature preprocessing is carried out, and the training set and testing set are divided according to a certain proportion. Thirdly, XGBoost is optimized by using ACO, then the optimized XGBoost algorithm is used to evaluate the importance of each feature. The obtained importance value will be used as threshold to obtain the optimal feature subset. Finally, the obtained feature subset will be detected and recognize by XGBoost classifier. Main steps of the proposed model are described in detail in the rest of Sect. 3.

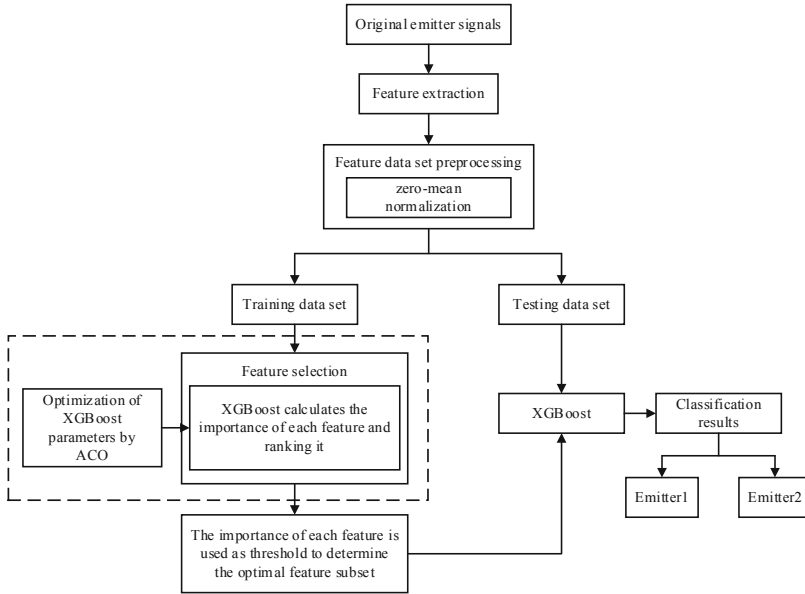


Fig. 1. The model of specific emitter identification based on ACO-XGBoost feature selection

3.1 Feature Extraction Based on Lifting Wavelet Package Transformation

Feature extraction is a key step of SEI. Feature extraction extracts the effective classification features in the original emitter signals through data conversion or data mapping, and obtains a new feature space. This step can reduce signal flow, directly affects the performance of classifiers and is an effective data compression method [15].

In view of the excellent time-frequency resolution and efficient operation ability of lifting wavelet package transformation, it is used as a tool for feature extraction in this paper. This method is helpful to expand feature set and even obtain more feature information through secondary change, so as to provide more choices for the final selection of relevant features.

Twelve statistical characteristic parameters, namely mean value, mean amplitude, square root amplitude, standard deviation, effective value, peak-to-peak value, shape factor, impulse factor, crest factor, skewness, kurtosis and clearance factor, are used. In addition, standardized relative energy is used, which is defined in reference [16].

Combined with lifting wavelet package decomposition and reconstruction, the characteristic parameter system is established as follows [17]:

Firstly, for a group of signals, the number of decomposition layers CS is limited to 5, and lifting wavelet packet tree is constructed. The optimal wavelet packet tree is obtained by using first order decomposition and then order search algorithm, and then the optimal wavelet packet tree is adjusted to a two-layer full binary tree.

Secondly, the information is concentrated on node (2, 0), (2, 1), (2, 2) and (2, 3) of the full binary tree. Twelve statistical characteristic parameters of four node's coefficient are

extracted respectively, and standardized relative energy which were mentioned above of four nodes are extracted.

Thirdly, single branch reconstruction is carried out for each node, and twelve statistical characteristic parameters are extracted from four single branch reconstruction signals respectively.

Finally, the second layer nodes are used to reconstruct the original signal and twelve statistical characteristic parameters of the reconstructed original signal are extracted.

According to the twelve statistical characteristic parameters of reconstructed original signal (label 1–12), each twelve statistical characteristic parameters of four node's coefficient of the second layer of wavelet packet decomposition (13–60), each twelve statistical characteristic parameters of four single branch reconstructed signals (61–108) and standardized relative energy of four nodes (109–112), the feature set $set = \{v_i | v_i = v_1, v_2, \dots, v_n\}$, $i = 1, 2, \dots, n = 112$ is constructed.

3.2 Feature Importance Scoring Based on XGBoost

XGBoost [18] is a kind of boosting algorithm. Based on GBDT algorithm, XGBoost algorithm carries out second-order Taylor expansion of loss function and adds a regular term, which avoids overfitting and effectively accelerates the convergence speed. By continuously adding new decision trees to fit the residual of previous prediction, XGBoost algorithm reduces the residual between the predicted value and the real value. Finally, the prediction accuracy is improved and the feature importance score is obtained. The prediction of XGBoost is described as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (5)$$

where K is the number of decision trees, f_k is the k th sub model, x_i is the i th input sample, F is the set of all decision trees.

The objective function of XGBoost consists of loss function and regular term:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (6)$$

where t is the number of iterations, l is a differentiable convex loss function that measures the difference between the prediction \hat{y}_i and the target y_i , $\hat{y}_i^{(t-1)}$ is the prediction of the previous $t - 1$ iteration, $\Omega(f_t)$ is the regular term of the k th iteration and its equation is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (7)$$

where γ and λ are regular term coefficients which can prevent the decision tree from being too complex, T is the number of leaf nodes, w is the leaf weight.

The feature importance score can be calculated as follows:

$$IS_i = \{x | x = w_i v_i\} \quad (8)$$

where v_i is the feature set $set = \{v_i | v_i = v_1, v_2, \dots, v_n\}$, $i = 1, 2, \dots, n$ and w_i is the weight of corresponding feature.

XGBoost algorithm carries out second-order Taylor expansion of loss function [19], then finds the minimum value of the objective function and calculates the corresponding optimal value by

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_j + \lambda} + \gamma T \quad (9)$$

where q represents the structure of each tree that maps an example to the corresponding leaf index, $I_j = \{i | q(x_i) = j\}$ is the instance set of leaf j , g_i is the first derivative of sample x_i , h_i is the second derivative of sample x_i . Equation (9) can be used to evaluate a tree structure. The smaller the value, the better the model.

The loss reduction which is also known as gain after split is shown as follows:

$$L_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_j + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_j + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_j + \lambda} \right] - \gamma \quad (10)$$

where the four terms on the right side of the equation represent the left subtree and right subtree score after splitting, the score before splitting, and the complexity regularization penalty coefficient respectively. In Eq. (10), $I = I_L \cup I_R$, when the reaches the depth limit or $L_{split} < 0$, the tree will stop splitting.

3.3 XGBoost Optimization Based on ACO

XGBoost has many parameters, and the selection of parameters affects the performance of the model. Reasonable parameters setting can significantly improve the accuracy of XGBoost. Some parameters are χ , δ , and ε which are discrete, and γ , λ and ω which are continuous. In this paper, ACO is used to optimize these parameters.

ACO is a swarm intelligence algorithm developed by natural genetics and evolution. It has great global search ability and can get positive feedback results. The basic idea of ACO is to use walking path of ants to represent the feasible solution, and all paths of the all ant colony constitute the solution space of the optimal problem. Ant with shorter paths releases more pheromone. With the advance of time, the pheromone on the shorter path gradually increase, and the number of ants choosing the path is also increased. Finally, the all ants will focus on the best path under the action of positive feedback and the best path is the optimal solution. The main steps of ACO are path selection and pheromone updating [20]. Path selection formula is shown as follows:

$$P_{ij}^k(t) = \frac{\{\max(\tau_i) - [\tau_{ij}(t-1)]^\alpha\} \eta_i^\beta}{\max(\tau_i) \sum_{e_{wj} \notin tabu_k} \eta_w^\beta}, \quad e_{ij} \notin tabu_k \quad (11)$$

where $P_{ij}^k(t)$ denotes the probability that ant k choose its path from node s_i to s_j at time t , $\tau_{ij}(t)$ is pheromone amounts, η_i is the heuristic function and the specific expression depends on the specific problem, α is the importance of pheromone amounts, reflecting

the impact extent of information for an ant to select a new path or not, β is the importance of heuristic function, e_{ij} is the edge from node s_i to s_j , taboo list $tabu_k$ records sides ant k have walked. In Eq. (11), $\tau_{ij}(t)$ is calculated as follow:

$$\tau_{ij}(t) = (1 - \rho)\tau_{ij}(t - 1) + Q\varphi'(tabu^t) \quad (12)$$

where ρ ($0 < \rho < 1$) is the pheromone evaporation coefficient, Q is a constant, which is determined according to ρ and adjusts the size of pheromone increment. $\varphi'(tabu^t)$ is the objective function value of the path $tabu^t$ at time t [21].

The ACO-XGBoost model is shown in algorithm 1.

Algorithm 1. ACO-XGBoost

- 1 **Input:** the parameters of ACO and the parameters to be optimized of XGBoost
 k : the number of ants; $\tau_{ij}(0)$: pheromone amounts before executing ACO;
 α : importance of pheromone amounts; β : importance of heuristic function;
 ρ ($0 < \rho < 1$) : pheromone evaporation coefficient; Q : a constant;
 K : the whole number of ants; N : iterations;
 the range of discrete parameters and continuous parameters namely C
 - 2 **Output:** optimal combination of parameters namely C'
 - 3 initialize continuous parameters and discrete parameters;
 - 4 **for** $i=1: N$ **do**
 - 5 **for** $j=1: K$ **do**
 - 6 ant k calculates the transition probability according to equation (11)
 - 7 **if** (selected edges do not satisfy constraints)
 - 8 {discard this selection; **break**}
 - 9 **if** (the iteration optimal solution is better than the current global optimal solution)
 - 10 {update the current global optimal solution; update the pheromone in this iteration according to equation (12)}
 - 11 **return** C' ;
-

In the training phase of ACO-XGBoost, the accuracy of XGBoost is used to evaluate the optimization method. Firstly, we initialize the parameters of ACO, set the value range of parameters which should be optimized in XGBoost. Then three discrete parameters are firstly optimized according to ACO and the results are assigned to XGBoost. In this condition, three continuous parameters are also optimized by ACO. Secondly, the test set is used to evaluate the model after training and determines whether the current parameter value is the optimal value. Finally, ACO-XGBoost follows these steps to find the best parameters after the specified number of iterations.

4 Experiment Procedure and Result Analysis

4.1 Datasets

The CPU is Inter Xeon E5-2630, memory size is 192GB, operating system is Ubuntu 16.04, programming environment is Python 3.8 of the experiment.

The experimental data comes from two emitters. The acquisition environment is basically a clean environment without noise. The signal data are obtained under 10 different acquisition states. The 10 specific signal parameters are shown in Table 1.

Table 1. Signal parameters

	Carrier wave/MHz	Modulation mode	Signal bandwidth/Hz	Sampling frequency/M	Interval time/ms
1	55	QPSK	25 k	1	20
2	75	QPSK	25 k	1	20
3	420	QPSK	2 M	20	2
4	420	QPSK	5 M	50	1
5	420	QPSK	10 M	80	1
6	420	QPSK	20 M	100	2
7	2000	QPSK	2 M	20	2
8	2000	QPSK	5 M	50	1
9	2000	QPSK	10 M	80	1
10	2000	QPSK	20 M	100	1

For each radio station, 200 groups (4096 data of a group) data are selected in each acquisition state (4000×4096 data in total), and selecting 75% (3000) for training, selecting 25% (1000) for testing. According to Sect. 3.1, after feature extraction, a feature set $= \{v_i | v_i = v_1, v_2, \dots, v_{112}\}$, $i = 1, 2, \dots, 112$ can be constructed. In this paper, the features are extracted from amplitude, channel I and channel Q respectively. And then a dataset named $dataset_{original} = \{v_i | v_i = v_1, v_2, \dots, v_{336}\}$, $i = 1, 2, \dots, 336$ is achieved. To verify the efficiency and effectiveness of the proposed method, Gaussian white noise is added to the original signal to make the signal-to-noise ratio be 10 dB and 5 dB respectively. The other two datasets are named $dataset_{10dB}$ and $dataset_{5dB}$.

4.2 Data Preprocessing

In order to unify the order of magnitude, increase comparability and speed up convergence of data samples, zero-mean normalization is used and the formula is shown as follows [14]:

$$v'_{tb} = \frac{v_{tb} - \bar{X}_{vt}}{\delta_{vt}} \quad (13)$$

where v_{tb} is the b th feature value of feature t , \bar{X}_{vt} is the mean value of feature t , δ_{vt} is the standard deviation of feature t .

4.3 Feature Selection by Using Important Value of XGBoost

Based on Eq. (8), the score map of feature importance in the original radio dataset (the top 20 largest is shown) is got as Fig. 2. In Fig. 2, f46 represents the 47th feature.

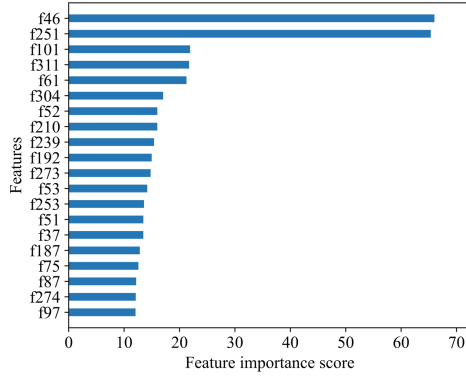


Fig. 2. Ranking of feature importance

To obtain the optimal feature subset, the importance of each feature is used as threshold to select features. For $dataset_{original}$, The threshold and the number of features is shown in Table 2.

Table 2. Feature selection accuracy of thresholds

Threshold	Numbers of features	A (%)
0.000	336	98
0.001	129	97.8
0.002	84	98.7
0.003	73	98
0.005	45	98
...
0.091	1	51

In Table 2, by setting each importance as the threshold, the features are divided into some subsets and the accuracy of the current subset is calculated. In the experiment, the feature importance is sorted from small to large, and then the sorted importance is used as every threshold to select the best feature subset.

It can be seen from Table 2 that the subset composed of the first 84 features with the highest importance has the highest accuracy. In some cases, adding features cannot improve the accuracy or even reduce the accuracy, so feature selection is rather necessary.

4.4 The Effect of Parameters on XGBoost

The parameters in XGBoost will affect the final prediction result of the model. In this section, the effect of different parameters on XGBoost are shown Fig. 3. The classification accuracy without parameter optimization in $dataset_{original}$ was 0.98.

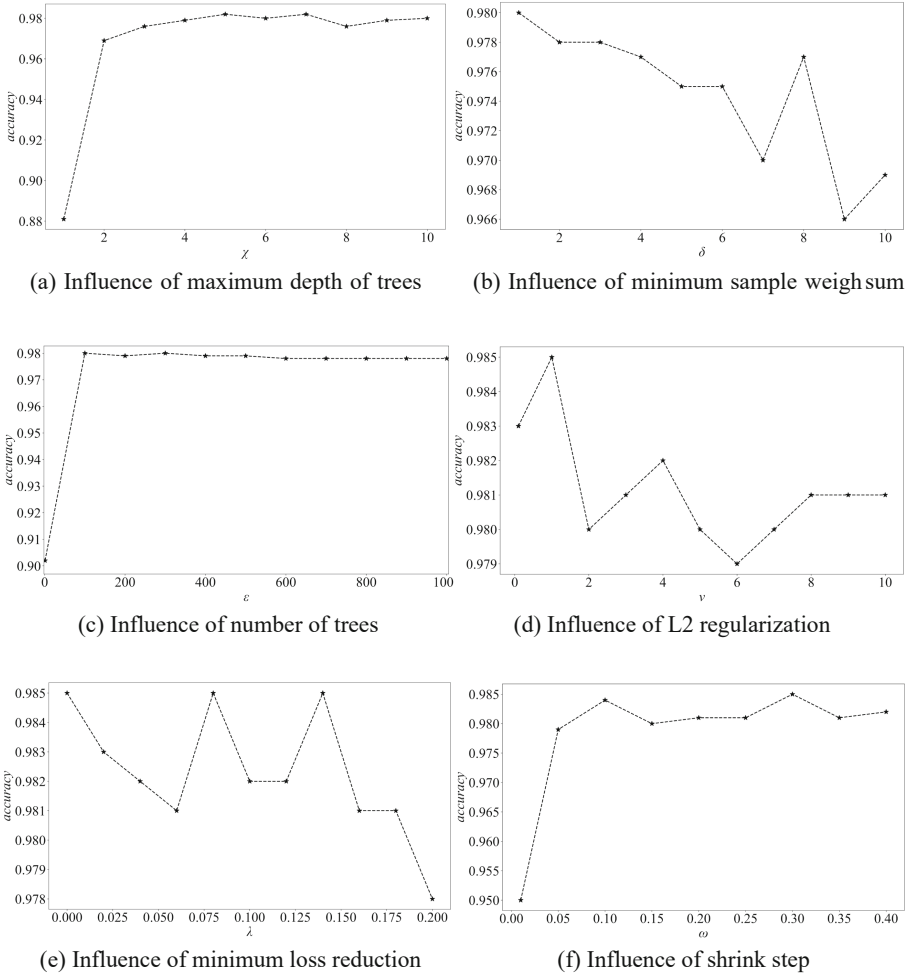


Fig. 3. Influence of XGBoost parameters on accuracy

Figure 3(a)(b)(c) are the change rule of prediction accuracy in the process to test the effect of three discrete parameters. As shown in Fig. 3(a), when $\chi = 5$, a higher accuracy can be obtained, and with the continuous increase of the tree depth, accuracy will not be further improved. The larger tree depth makes the model more complex and lead to overfitting easily, which will in turn reduce the prediction accuracy of test set. Figure 3(b) shows that the accuracy of the model basically decreases with the increase

of δ , which indicates that the larger the parameter is, the less easy it is to overfit, but it may result in underfit. It can be seen from Fig. 3(c) that when the value of ε is between 100 and 300, a higher accuracy can be obtained. However, with the continuous increase of the number of trees, the change range of accuracy is small and the model will be easy to overfit.

For three continuous parameters, several values within the corresponding value range are selected and the prediction accuracy of each value is calculated. Then line charts are drawn. Figure 3(d)(e)(f) show the influence and change law of three continuous parameters on the accuracy of SEI. In Fig. 3(d) and Fig. 3(e), when the values of γ and λ are relatively small, the accuracy has been improved. With the increase of parameter value, although the accuracy has been fluctuating, a better accuracy value has not obtained. Figure 3(f) shows that a relatively highest accuracy can be obtained when ω is around 0.3. The smaller the learning rate is, the smaller the calculation speed is. If the learning rate is too large, it may not converge.

By changing the parameters of XGBoost, the accuracy of SEI is greatly affected, which shows that the most appropriate value of parameters will maximize the use of all characteristics of signal data and obtain the optimal model. In this section, the influence of various parameters on XGBoost has been explored. In order to obtain the optimal value of these parameters, ACO is used to further optimize the parameters.

4.5 XGBOOST's Parameter Optimization Based on ACO

This section aims to use ACO to optimize XGBoost and the variation curve of the predicted value with the iterations of ACO is shown in Fig. 4.

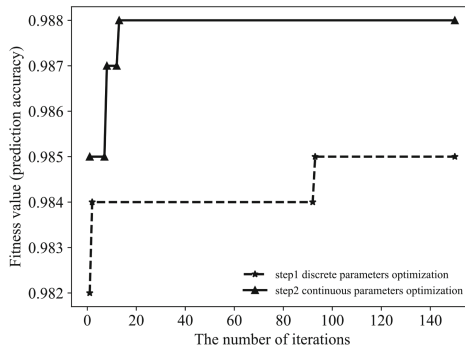


Fig. 4. Iteration curves of XGBoost optimized by ACO

In Fig. 4, the range we set and the default value of six parameters which have mentioned above are all shown in Table 3. Initializing $k = 45$, $\tau_{ij}(0) = 0$, $\alpha = 1$, $\beta = 0$, $\rho = 0.8$, $Q = 1$, $N = 150$. In Fig. 4, firstly, ACO is used to optimize three discrete parameters. Secondly, we optimize the three continuous parameters when we have determined and brought the discrete parameters into XGBoost model. Compared with default parameters of XGBoost, the accuracy is improved by 0.80% after two steps

of optimization. Under the given termination condition, the parameters optimized by ACO are shown in Table 3.

Table 3. Model parameters

Model parameters	Parameter description	Range we set	Default	ACO-XGBoost
χ	Maximum depth of trees	[1, 10]	6	5
δ	Minimum sample weigh sum in leaf node	[1, 10]	1	1
ε	Number of trees	[1,1000]	100	202
γ	L2 regularization penalty coefficient	[0.1,10]	1	5.91199579
λ	Minimum loss reduction	[0,0.2]	0	0
ω	Shrink step	[0.01,0.4]	0.3	0.24792934

4.6 Performance Comparison

Based on the above introduction, datasets and evaluation criteria, the classification prediction experiment is designed, and the method proposed in this paper is compared with DT, RF, GBDT and XGBoost with default parameters. Due to certain randomness, we conduct three experiments of DT, RF and GBDT respectively and take the average value. The results are shown in Table 4.

Table 4. Performance comparison of different feature selection models

Dataset and evaluation criteria		DT	RF	GBDT	XGBoost	ACO_XGBoost
<i>dataset_{original}</i>	A (%)	95.37	98.10	97.80	<u>98.70</u>	98.90
	q	21	68	49	84	<u>39</u>
<i>dataset_{10dB}</i>	A (%)	66.07	72.27	<u>72.40</u>	72.30	73.00
	q	49	192	11	46	<u>35</u>
<i>dataset_{5dB}</i>	A (%)	55.90	59.10	58.40	<u>59.20</u>	59.30
	q	92	141	179	<u>40</u>	35

In Table 4, the bold value is the optimal value and the underlined value is the sub-optimal value. It can be seen that ACO-XGBoost method proposed in this paper can effectively improve the prediction ability and obtain a relatively small feature subset of most datasets. In *dataset_{original}*, accuracy of ACO_XGBoost has been improved by 0.20%-3.53% compared with the other four comparison algorithms and q is the second

smallest. In $dataset_{10dB}$, accuracy index has been improved by 0.60%-6.93%. Also, q is the second smallest. In $dataset_{5dB}$, accuracy increased by 0.10%-3.40%, and the number of selected features q is smallest. Experiment shows that the proposed method in this paper not only obtains the optimal classification results but also acquired the feature subsets with relatively small feature number. A balance is got between the classification accuracy and the number of selected numbers.

5 Conclusion

In this paper, we present a new model of SEI based on ACO-XGBoost feature selection to improve the identification ability of specific emitter. Our method consists of the following three steps: 1) Lifting wavelet packet decomposition and deconstruction are used for feature extraction and then a characteristic parameter system is constructed. 2) ACO is used to optimize the parameters of XGBoost. The optimization of discrete parameters is carried out firstly and the results are brought into the process of continuous parameter optimization. 3) The optimized XGBoost model is used for feature selection. The importance of each feature is calculated, and all the importance values are used as thresholds to select the optimal features to form the optimal feature subset. Experimental results on three datasets show that our proposed method can keep a balance between accuracy and selected features. In our future work, we will further control the number of selected features without reducing the prediction accuracy.

Acknowledgement. This paper is supported by the National Natural Science Foundation of China (71901215, 61371196), China Postdoctoral Science Foundation Funded Project (20090461425, 201003797), and the Nation Important Scientific and Technological Special Project (2015ZX01040201-003).

References

1. Satija, U., Trivedi, N., Biswal, G., Ramkumar, B.: Specific emitter identification based on variational mode decomposition and spectral features in single hop and relaying scenarios. In: IEEE Transactions on Information Forensics and Security, pp. 581–591. (2019)
2. Han, J., Zhang, T., Wang, H., Ren, D.: Communication emitter individual identification based on 3D-Hibert energy spectrum and multi-scale fractal features. In: Journal on Communications **38**(4), 99–109 (2017, in Chinese). <https://doi.org/10.11959/j.issn.1000-436x.2017080>
3. Xiu, Y., Zhao, S., Chen, H., Li, C.: I-mRMR: incremental max-relevance, and min-redundancy feature selection. In: Shao, J., Yiu, M.L., Toyoda, M., Zhang, D., Wang, W., Cui, B. (eds.) APWeb-WAIM 2019. LNCS, vol. 11642, pp. 103–110. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26075-0_8
4. Xue, H., Jiang, J., Shao, Y., Cui, B.: FeatureBand: a feature selection method by combining early stopping and genetic local search. In: Shao, J., Yiu, M.L., Toyoda, M., Zhang, D., Wang, W., Cui, B. (eds.) APWeb-WAIM 2019. LNCS, vol. 11642, pp. 27–41. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26075-0_3

5. Zhang, W., Chen, C., Jiang, L., Bai, X.: A new feature selection algorithm based on category difference for text categorization. In: Shao, J., Yiu, M.L., Toyoda, M., Zhang, D., Wang, W., Cui, B. (eds.) APWeb-WAIM 2019. LNCS, vol. 11642, pp. 322–336. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26075-0_25
6. Wang, J., Wei, J., Yang, Z.: Supervised feature selection by preserving class correlation. In: CIKM 2016, pp. 1613–1622. ACM, (2016). <https://dx.doi.org/https://doi.org/10.1145/2983323.2983762>
7. Rostami, M., Berahmand, K., Nasiri, E., Forouzandeh, S.: Review of swarm intelligence-based feature selection methods. *Engineering Appl. Artificial Intelligence* **100**, 104210 (2021)
8. Wang, T., Wang, X., Liu, J.: Prediction of moisture content of *Aquilaria Sinensis* leaves based on RFE_RF algorithm model. In: *Journal of Nanjing Forestry University (Natural Sciences Edition) [J/OL]* (2022, in Chinese)
9. Chen, Z., Lu, N.: Network intrusion detection model based on Random Forest and XGBoost. In: *Journal of Signal Processing* **36**(7), 1055–1064 (2020, in Chinese)
10. Zhou, J., He, P., Qiu, R., Chen, G., Wu, W.: Research on intrusion detection based on random forest and gradient boosting tree. *Journal of Software* **32**(10), 3254–3265 (2021, in Chinese)
11. Feng, J., Liang, J., Qiang, Z., Li, X., Chen, Q., Liu, G., et al.: Effective techniques for intelligent cardiocography interpretation using XGB-RF feature selection and stacking fusion. In: *International Conference on Bioinformatics and Biomedicine*, pp. 2667–2673 (2021)
12. Gong, H., Zhang, H., Zhou, L., Liu, Y.: An interpretable artificial intelligence model of Chinese medicine treatment based on XGBoost algorithm. In: *Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 1550–1554 (2020)
13. Liu, L., Tang, S., Wu, F., Wang, Y., Wang, J.: An ensemble hybrid feature selection method for neuropsychiatric disorder classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **19**(3), 14591471 (2021). <https://doi.org/10.1109/TCBB.3021.3053181>
14. Qiao, N., Li, Z., Zhao, G.: Intrusion detection model of internet of things based on XGBoost-RF. In: *Journal of Chinese Computer System* **43**(1), 152–158 (2022, in Chinese)
15. Liu, M., Yang, Z., Zhang, J.: Specific emitter identification method for aerial target. In: *Systems Engineering and Electronics* **41**(11), 2408–2415 (2019, in Chinese). [https://doi.org/10.16798/j.issn.1001-506X\(2019\)11-2408-08](https://doi.org/10.16798/j.issn.1001-506X(2019)11-2408-08)
16. Cao, J., Zhang, P., Zhang, Y.: Feature extraction of an engine cylinder head vibration signal based on lifting wavelet package transformation. *Journal of Vibration and Shook* **27**(2), 34–37 (2008, in Chinese)
17. Cao, J., Zhang, P., Ren, G.: Feature selection of vibration signal based on ant colony optimization. *J. Vibration and Shock* **27**(5), 24–31 (2008, in Chinese)
18. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. San Francisco (2016)
19. Li, L., et al.: Spectroscopy-based food internal quality evaluation with XGBoost algorithm. In: U, L.H., Xie, H. (eds.) APWeb-WAIM 2018. LNCS, vol. 11268, pp. 56–64. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01298-4_6
20. Yang, L., Li, K., Zhang, W., Ke, Z., Xiao, K., Du, Z.: An improved chaotic ACO clustering algorithm. In: *Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications*, pp. 1642–1649 (2018)
21. Mu, N., Xu, X., Zhang, X.: Ant colony optimization based salient object detection for weak light images. In: *Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence, Internet of People and Smart City Innovations*, pp. 1432–1437 (2018)