Communication signal classification and recognition method based on GA-LSSVM classifier

QingChuan Meng1,2*, Xiaoming Chen2, YuLin Zhu1, DaYong Hu1, Rui Li1, YongTiao Jiang1
1 HeiLongJiang Province Cyberspace Research Center, Harbin, HeiLongJiang, 150090, China
2 College of Computer Science And Technology, Harbin Engineering University, Harbin, HeiLongJiang, 150090, China
*Corresponding author’s e-mail: mqchappy365@163.com

Abstract. In order to recognize signals more efficiently and accurately, this paper proposes a method of communication signal classification and recognition based on GA-LSSVM classifier. Firstly, GA algorithm is used to optimize the penalty factor and kernel function parameters of the two main parameters in the model of support vector machine classifier. By constructing a GA-LSSVM classifier for wireless communication signals, combining with the characteristic parameters of modulated signals, the signals are identified and simulated under different SNR conditions, and the analysis and verification are carried out. The performance of GA-LSSVM classifier is studied. The simulation results show that GA-LSSVM classifier has better recognition and classification performance than other classifiers in different SNR environments, and more than 90% accuracy can be achieved when SNR is greater than 0 dB.

1. Introduction
Nowadays, computer communication technology, which combines computer technology with wireless communication technology, is widely used in military, civil and other fields, and plays an important role in modern communication. Comparing with traditional communication technology, computer wireless communication technology has strong communication ability, large amount of information and adapts to space-time conversion. It has broad prospects in realizing intelligent communication, real-time detection and other applications. In the computer wireless communication system, combined with computer technology, the detection of wireless communication signals and automatic identification of their modulation types are the current research hotspots, and the research of automatic identification of wireless communication signals based on computer technology is of great significance.

In order to realize rapid and accurate recognition of wireless communication signals, some researchers have proposed a feature extraction method based on computer technology such as instantaneous phase and frequency of signals, wavelet transform, Hilbert Huang transform and so on[1]. For the design of computer classifier, in order to improve the efficiency of modulation recognition and improve the efficiency of modulation recognition[2]. In this paper, a new wireless communication signal classifier is constructed based on a computer modulation recognition technology of GA-LSSVM to make the recognition and classification of modulation signals more accurate and efficient.
2. pattern recognition

2.1 Genetic Algorithms
Genetic algorithm (GA) is an evolutionary algorithm. Its basic principle is to code the parameters of the problem, select, cross and mutate the chromosomes by iteration, and optimize the selection of the target parameters by exchanging different chromosome information, following the principle of "natural selection and survival of the fittest" in nature[3]. The minimum support vector machine parameters are optimized by genetic algorithm[4-5]. The basic steps of the algorithm are as follows:

(1) Randomly generate 100 individuals \( s_1, s_2, \ldots, s_{100} \) to form the initial population \( S \) in the search space, setting the algebra counter \( t = 1 \);
(2) Calculating the fitness \( f_i = f(s_i) \) of each individual \( s_i \) in \( S \);
(3) Calculating probability.

\[
p(s_i) = \frac{f_i}{\sum_{j=1}^{100} f_j}, \quad i = 1, 2, \ldots, 100
\]

According to the above probability distribution equation, individuals are randomly selected from \( S \) in sequence and chromosomes are copied 100 times, and new chromosome groups are used instead of the original group to obtain a new group \( S_2 \).
(4) According to the crossover rate \( P_c \), the number of chromosomes \( C \) participating in the intersection is selected, and \( C \) chromosome pairs are randomly selected from \( S_2 \) for cross operation, and the new chromosome is used instead of the original chromosome population to obtain the population \( S_3 \).
(5) According to the mutation rate \( P_m \) and the number \( m \) of mutations, \( m \) chromosomes are randomly selected from \( S_3 \) for mutation operation, and a new chromosome is used instead of the original chromosome population to obtain a population \( S_4 \).
(6) Repeat the above steps until the optimal solution is obtained.

2.2 Support Vector Machine
The least squares support vector machine (LSSVM) is mainly for the two-class problem. The basic idea is to use the principle of structural risk minimization to construct a classification plane that satisfies the condition. By using the secondary plane as the decision surface, the classification is correct and the two types of samples are made. Maximize the spacing between them. LSSVM has the following advantages:

(1) Effectiveness: It is one of the best ways to solve practical problems;
(2) Robustness: no fine tuning is required;
(3) Simple calculation: the method realizes the use of simple optimization techniques;
(4) Theoretical perfection: Based on the framework of VC generalization theory.

The concept of inner product kernel between support vector machine \( x(i) \) and vector \( x \) extracted by input space is the key to the algorithm. The architecture is shown in Figure 1.
Input \[x(1)\] \[x(2)\] ... \[x(3)\]. Bias \[b_k(x - x_1)\] \[b_k(x - x_2)\] \[b_k(x - x_m)\].

Output

Figure 1. Support vector machine architecture

K is a kernel function, and the complexity of the vector machine differs depending on the kernel function. The kernel functions that are often used are the following:

1. Polynomial kernel function
   \[K(x, x_j) = [(x, x_j) + 1]^q\]

2. Radial basis kernel function
   \[K(x, x_j) = \exp\left(-\frac{\|x - x_j\|^2}{2\delta^2}\right)\]

3. Multi-layer neural network tangent hyperbolic kernel function
   \[K(x, x_j) = \tanh(v(x \cdot x_j) + c)\]

In equation (1), equation (2) and equation (3):

\[Q = -I \sum_{j=1}^{N} W_j \begin{bmatrix} 1 & (x_j, x_j) - |A_{xj} - I_j| \end{bmatrix} \]

In equation (3), \(q, \delta, v,\) and \(c\) are all kernel function parameters. The kernel function is widely used in many fields such as target prediction.

This paper sets \(D = \{(x_k, y_k) | k = 1, 2, ..., N\}\). Where \(x_k \in R\) is the value of the corresponding competition when the optimal radius is selected for the different position tasks, and \(y_k \in R\) is the lowest price corresponding to the optimal radius. In the weight space \(\omega\), the least squares support vector machine solves the objective function as follows:

\[\min_{w, b, e} \phi(a, b, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^{N} e_k^2\]

The constraint is

\[y_k \left[\omega^T \phi(x_k) + b\right] = 1 - e_k, \quad k = 1, 2, ..., N\]

Define the Lagrang function as:

\[L(w, b, e, \alpha) = \phi(w, b, e) - \alpha_k \sum_{k=1}^{N} \{\alpha_k y_k \left[\omega^T \phi(x_k) + b\right] - 1 - e_k\}\]

In the formula, the Lagrangian multiplier \(\alpha_k \in R\). The equation (8) is optimized, that is, the partial derivatives of \(w, b, e_k\), and \(\alpha_k\) are obtained, and the result is equal to 0, and finally the matrix equation can be obtained as follows:

\[\begin{bmatrix} 0 & -Y^T \end{bmatrix} \begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}\]

In equation (9):


\[
Z = \begin{bmatrix} \phi(x_1)^T y_1, \phi(x_2)^T y_2, \ldots, \phi(x_N)^T y_N \end{bmatrix} \\
Y = \begin{bmatrix} y_1, y_2, \ldots, y_N \end{bmatrix} \]

\[I = [1, 1, \ldots, 1]\]  
\[e = [e_1, e_2, \ldots, e_N]^T\]  
\[\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]\]

Bring the Mercer condition into \(\Omega = ZZ^T\)

\[
\Omega_M = y_k, y_i \phi(x_k) = y_k, y_i \phi(x_k, x_i)
\]

Further, the solution of the equation (7) and (8) can be realized. In addition, this paper selects the decision function of the least squares support vector machine:

\[
y(x) = \text{sgn} \left[ \sum_{k=1}^{N} \alpha_k y_k \phi(x_k, x_i) \right]
\]

### 3. Digital Signal Modulation Recognition Model

#### 3.1 Feature Parameter Extraction

When the communication signal is identified, the feature parameter extraction by the sample signal to be identified can effectively reduce the influence of the data amount on the signal recognition. By converting the high-dimensional sample data into the low-dimensional space, the spatial dimension of the data can be effectively reduced. The selection methods of the characteristic parameters are different, the estimation methods are different, and the performance of the classification algorithm is also different. The evaluation of different characteristic parameters can be carried out from the following aspects: whether the characteristic parameters significantly characterize the characteristics of the modulated signal, and whether the characteristic parameters can effectively suppress the noise interference. In this paper, different feature parameters are selected from the spectrum, signal complexity and instantaneous information of the signal to be identified to form the feature set of the sample signal.

1. **Zero-center normalized homeopathic amplitude spectral density maximum** \(\gamma_{\text{max}}\):

\[
\gamma_{\text{max}} = \max \left[ \text{FFT} \left[ \frac{a_{\text{cn}}(i)}{N_s} \right] \right]
\]

\[
m = \frac{1}{N} \sum_{i=1}^{N} a(i)
\]

\[
a_{\text{m}}(i) = a(i) / m
\]

\[
a_{\text{acn}}(i) = a_{\text{m}}(i) - 1
\]

Where \(N_s\) represents the number of signal sampling points, \(a(i)\) is the instantaneous amplitude of the signal, \(m_s\) is the instantaneous amplitude averaging, \(a_{\text{m}}(i)\) is the instantaneous amplitude after normalization, and \(a_{\text{acn}}(i)\) is the instantaneous amplitude after centralized normalization. By normalizing the instantaneous amplitude, the interference of the channel gain on the characteristic parameters can be eliminated. The characteristic parameter of the signal \(\gamma_{\text{max}}\) characterizes the amplitude change information and can be used to distinguish between (FSK, 4PSK) and (ASK, 2PSK, 16QAM).

2. **Zero center normalized instantaneous amplitude absolute value standard deviation**:

\[
\delta_{\text{m}} = \left[ \frac{1}{N_s} \sum_{j=1}^{N_s} a_{\text{m}}^2(i) \right] \left( \frac{1}{N_s} \sum_{j=1}^{N_s} a_{\text{m}}^2(i) \right)^{1/2}
\]
In the process of modulation of 2PSK signal, the amplitude is not completely symmetrical. The characteristic parameter delta AA represents the change degree of absolute value of signal amplitude, and can distinguish 2ASK and 2PSK signals.

(3) Standard deviation of amplitude envelope $E$:

$$\delta_{aa} = \left( \frac{1}{N_s} \sum_{i=1}^{N_s} \left[ a(i) - \frac{1}{N_s} \sum_{i=1}^{N_s} a(i) \right] \right)^{1/2}$$

(15)

The 4PSK and 16QAM carrier amplitudes have more values, and have larger instantaneous amplitude fluctuations and amplitude envelope variances than 2ASK and 2PSK signals. The characteristic parameter $E$ is used to describe the degree of fluctuation of the instantaneous amplitude of the signal, which can effectively identify 2ASK, 4ASK, 2PSK and 16QAM.

(4) Characteristic parameters $R$:

$$R = \frac{\mu}{\sigma^2}$$

(16)

Where $\mu$ is the average value of the signal amplitude envelope, $\sigma$ is the square standard deviation of the signal amplitude envelope. The characteristic parameters reflect the fluctuation of the signal amplitude envelope, distinguishing between 16QAM and 4ASK.

(5) Normalized instantaneous frequency absolute standard deviation $\sigma_{af}$:

$$\sigma_{af} = \left[ \frac{1}{C} \left( \sum_{a_i > a_t} f_3^2(i) \right) - \frac{1}{C} \sum_{a_i > a_t} |f_3(i)|^2 \right]^{1/2}$$

(17)

In the equation, $a_t$ denotes the threshold of signal amplitude, which is often used to discriminate non-weak signal segments with a value of 1; $C$ denotes the number of samples in non-weak signal segments at points; and $f_3(i)$ denotes the instantaneous frequency after normalization of non-weak signal segments. The characteristic parameter $\sigma_{af}$ characterizes the absolute frequency information state in the signal, and can be used to discriminate 2FSK and 4FSK.

(6) Box dimension $D_B$ of signal

$$D_B = 10 \cdot \log_2 \frac{d(\Delta)}{d(2\Delta)}$$

$$d(\Delta) = \sum_{i=1}^{N_s} |s(i) - s(i+1)|$$

$$d(2\Delta) = \sum_{i=1}^{N_s/2} [\max[s(2i-1),s(2i)],s(2i),s(2i+1)] - \min[s(2i-1),s(2i),s(2i+1)]$$

(18)

Box dimension originates from fractal theory, which can accurately describe the irregularity, complexity and geometric size of signals. Digital modulation signal is a regular sequence which changes with time. The characteristic parameters can be used to distinguish 4PSK from 2FSK and 4FSK.

### 3.2 Modulation Recognition Based on GA-LSSVM

GA-LSSVM classifier is used to recognize and classify the feature set of sample signals for modulation recognition. The specific display steps are as follows.

Step 1: According to the feature extraction formula, the relevant feature parameters of seven kinds of digital modulation signals are extracted, and the feature set of modulation recognition sample signals is selected as input data of LSSVM classifier.

Step 2: Choose the appropriate kernel function.

Step 3: Genetic algorithm is used to optimize the parameters $C$ and of SVM classifier. After finding the optimal solution, the SVM model is brought in to form the GA-LSSVM classifier.
Step 4: Input the training sample data into the GA-LSSVM classifier, realize the training process of the classifier, locate the optimal classification hyperplane, and construct the GA-LSSVM classification model. Bring the sample data into the classification model completed by the training, the output results are the types of modulation signals.

4. Simulation results and analysis
Among the signals with distributed Gauss noise, 2ASK, 4ASK, 2FSK, 4FSK, 2PSK, 4PSK and 16QAM are selected to wait for the recognition signals, and the corresponding characteristic parameters are extracted to construct the sample set. The GA-LSSVM classification model is used to identify and classify different digital signals.

The parameters of GA algorithm are as follows: search space range [0.001, 1000], population size is 100, iteration times are 100, crossover probability is 0.4, mutation probability is 0.2. To further verify the recognition performance of the proposed algorithm, the test results of different algorithms are shown in Figure 2.

![Figure 2. Accuracy comparison of different algorithms](image)

It can be seen that under the condition of the kernel function is the same, compare the test results with the results obtained by using particle swarm optimization algorithm, extending the swarm optimization algorithm and least squares support vector machine. It can be seen that at each signal-to-noise ratio point of founder, the classification error of least-squares support vector machine based on genetic algorithm is less than 10%, and its classification accuracy reaches more than 90%, showing obvious classification advantages

5. Conclusion
A new computer wireless communication classifier is constructed, which provides a new method for modulation recognition of radio digital signals. Firstly, the penalty factor and kernel function parameters of SVM classification are optimized by GA-LSSVM algorithm to improve the efficiency of classifier, and the performance of GA-LSSVM recognition classifier is evaluated under different signal and noise conditions. The simulation results of seven digital modulation signals show that the performance of GA-LSSVM classifier is better than other classifiers when the noise environment is different, and the classification error of the classifier can be controlled below 10% when the signal-to-noise ratio is greater than 0 dB.

References
[1] Wu, L. (2016) Research on Communication Signal Modulation Recognition Technology for Advanced Signal Processing Methods. J. Digital Communication World, 10:56.
[2] Wang, Z. (2016) Research on Communication Signal Modulation Recognition Method. J. Science and Technology Innovation, 29:69.
[3] Wang, Y., Duan, D., Cheng, H. (2016) Study on Modulation Recognition of Low Voltage Power Line Carrier Communication Signal Based on Neural Network. J. Electrical Measurement & Instrumentation, 53(11):51-55.

[4] Long, X., Zhang, H., Zhang, M. (2017) Modulation Recognition Algorithm for Wireless Communication Signals Based on Harmonic Average Fractal Box Dimension. J. Journal of Jiangsu University(Nat Sci Ed), 38(03):308-312.

[5] Wang, M., Cai, M. (2015) Modulation The Modulation Recognition of Digital Signals Based on Improved Genetic Algorithm. J. Digital Communication World, 02:65-68.