Radar signal recognition method based on SVM model

Letian Tian, Zhaowen Zeng*, Zhiyang Chang, Sheng Zhang
College of Information and Communication, National University of Defence Technology, Wuhan, Hubei, China

ABSTRACT

Radar signal recognition is a key step of electronic reconnaissance. In order to use various parameters of radar signal to identify its country in complex electromagnetic environment, this paper analyses the characteristics of radar signal and the development status of related recognition methods. Aiming at the support vector machine (SVM) classification algorithm in machine learning, this paper studies the feature selection and the establishment, evaluation and optimization of model, and completes the programming and application of the algorithm, the recognition accuracy and efficiency are improved. The research shows that when SVM is applied to radar signal recognition, the classification accuracy is more than 93% and the AUC value is more than 0.99.

Keywords: Radar signal, classification, identification, SVM

1. INTRODUCTION

Under the current information war conditions, the electromagnetic environment is getting worse and worse. As an information reconnaissance means, radar signal recognition plays a more and more important role in reconnaissance and early warning and assisting battlefield decision-making[1]. The traditional recognition method mainly matches the intercepted radar signal data with the existing information base, so as to determine the relevant characteristics of the radar signal to be recognized. This method relies too much on prior knowledge and has no learning ability, so it will be helpless when encountering strange radar signals[2]. As an intelligent method of analysing data, the most prominent feature of machine learning is its strong self-learning ability. Its excellent performance in data mining and intelligent recognition provides technical support for radar signal recognition[3].

In recent years, machine learning as an intelligent method of radar signal recognition, has become a research hotspot at home and abroad. Chen et al. proposed a recognition method based on artificial neural network, which realizes radar signal recognition through nonlinear mapping[4]. It has strong adaptive ability, but the process of determining the structure of artificial neural network is complex and easy to lead to over fitting. According to the rules of decision tree generation, Zhu et al. established C5.0 algorithm for radar signal recognition[5], which is simple and easy to use, and does not rely on prior knowledge, but it needs data preprocessing and parameter optimization, which is not ideal in decision efficiency. SVM has low dependence on prior experience and high classification accuracy, which has unique advantages in dealing with small samples and nonlinear data[6]. Aiming at the characteristics of many characteristic variables, complex data types and independent variables of radar signal data, this paper studies the application of SVM model in radar signal recognition.

2. THEORETICAL BASIS OF SVM

2.1 Basic principles

The basic principle of SVM is to project the N data into the k-dimensional space to form N points, in which the spatial dimension k is equal to the number of characteristic variables, and then find a hyperplane in the k-dimensional space, so that the N data points can be separated into two categories, and have the best reliability at the same time. Such a hyperplane is called the best hyperplane. When predicting the test set data, it can be judged and identified only according to the position relationship between it and the hyperplane.

2.2 Classification basis

An example of a binary classification problem in the two-dimensional space shown in Figure 1 is given. In this case, the hyperplane is a decision line.

* 1251463841@qq.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125062K
 · 0277-786X · doi: 10.1117/12.2661773

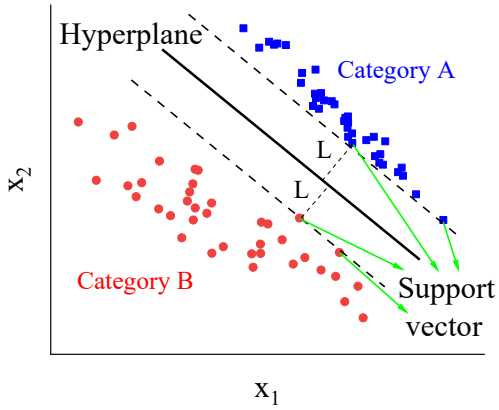

Figure 1. Two-dimensional spatial classification of SVM.

The point closest to the best decision line is called the support vector, and the line parallel to it and passing through the support vector is called the baseline. The best decision line should not only be able to accurately separate the data, but also be as far away from the support vector as possible, so as to ensure a certain degree of tolerance for new data. Therefore, the problem of finding the best decision line is transformed into the problem of solving the maximum interval between two baselines. The equation of the decision line can be set as

$$k_1x_1 + k_2x_2 + b = 0 \tag{1}$$

Then the baselines equation is

$$k_1x_1 + k_2x_2 + b = \pm 1 \tag{2}$$

Because the support vector is closest to the best decision line, there are constraints for all data points,

$$y(k_1x_{1i} + k_2x_{2i} + b) \geq 1,\ y = \pm 1 \tag{3}$$

Therefore, the maximum spacing distance can be obtained as

$$2L = \frac{2}{\sqrt{k_1^2+k_2^2}} \tag{4}$$

Because there are many characteristic variables in radar signal data, the classification problem rises to multi-dimensional space, which is represented by matrix

$$K = (k_1 \quad k_2 \quad \cdots \quad k_k),\ X = (x_1 \quad x_2 \quad \cdots \quad x_k) \tag{5}$$

The result is

$$2L = \frac{2}{\|K\|}, y(K \cdot X_i^T + b) - 1 \geq 0 \tag{6}$$

2.3 Solution of optimal hyperplane

Lagrange multiplier method, as a method to solve the extreme value of objective function under multiple constraints, can play a great role here. Therefore, Lagrange function is constructed

$$\mathrm{F(K, b, \lambda_i, c_i)} = \mathrm{f(K)} - \sum_{i=1}^{N} \lambda_i \varphi(\mathrm{K, b, c_i}) = \frac{\|K\|^2}{2} - \sum_{i=1}^{N} \lambda_i \left[y(\mathrm{K \cdot X_i^T + b}) - 1 - c_i^2\right] \tag{7}$$

Derivation of each parameter of F is obtained and the value of the derivative function is set to be 0, and then the results are substituted into equation (7) to get

$$F(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j y_i y_j\ (X_i \cdot X_j^T) \tag{8}$$

If the data points of different categories are staggered, the two categories cannot be separated, then the spatial conversion is needed to achieve classification, as shown in Figure 2.

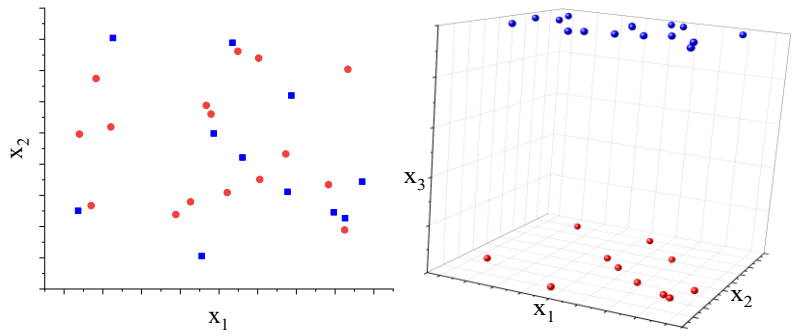

Figure 2. Classification by spatical conversation.

In order to simplify the operation, the Kernel Function is used to avoid the dimension conversion process. Kernel Function is very important for the prediction results and the performance of the model. There are four kinds of Kernel Function: poly, sigmoid, rbf and linear.

3. DATA REPROCESSING AND FEATURE SELECTION

Data preprocessing and feature selection are important prerequisites for data modeling and analysis. High-quality data and high-value features are the guarantee for the model to obtain excellent prediction results. The radar signal data has many characteristic variables, including numeric and categorical variables, and it has text data that cannot be processed by SVM algorithm, as shown in Table 1.

Table 1. Statistical analysis results of radar signal data.

Variable name	Variable type	Species	Variable code
Country	Categorical	2	y
Platform type	Categorical	3	x_1
Platform model	Categorical	31	x_2
Radar model	Categorical	30	x_3
Carrier frequency type	Categorical	2	x_4
Recurrence type	Categorical	5	x_5
Pulse width type	Categorical	4	x_6
Scan type	Categorical	5	x_7
Scan cycle	Categorical	38	x_8
Recurrence	Numerical	283	x_9
Pulse width	Numerical	230	x_{10}
Carrier frequency	Numerical	226	x_{11}
Discovery time	Numerical	450	x_{12}
Disappearance time	Numerical	450	x_{13}
Discovery orientation	Numerical	450	x_{14}
Disappearance orientation	Numerical	450	x_{15}
Total number of variables	16		
Sample total	450		

3.1 Data reprocessing

The purpose of data preprocessing is to enable it to be identified and learned by the algorithm. For the categorical variables containing text data, they are usually numericalized, so that the categorical variables are corresponding to the data set. The specific situation is shown in Table 2.

Table 2. Numerical results.

Variable code	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Numerical results	0, 1	1, 2, 3	1-31	1-30	0, 1	1-5	1-4	1-5	1-38

3.2 Feature selection

Feature selection refers to correlation analysis to filter variables with strong correlation with the target variable, which can effectively eliminate irrelevant noise in the data, reduce the complexity of the model, and prevent the model from overfitting[7].

3.2.1 Correlation Analysis between Typed Variables. Chi-square test is a common correlation test method in statistical analysis, which can quantitatively determine the correlation between categorical variables. The principle is to obtain the uncorrelated probability P by calculating chi square statistics[8]. The correlation between target variable and categorical variables is shown in Table 3.

Table 3. Correlation between categorical variables.

Variable code	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Chi-square statistic	21.34	450	450	2.38	90.86	4.76	137.50	33.04
Uncorrelated probability	0	0	0	0.12	0	0.19	0	0.32

3.2.2 Correlation Analysis of Categorical Variables and Numeric Variables. Univariate ANOVA is to determine the degree of influence of variables on the target by calculating the sample mean, and to measure the correlation between typed variables and continuous variables by correlation coefficients[9]. The results are shown in Table 4.

Table 4. Correlation between categorical and numerical variables.

Variable code	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
Degree of correction	1	1	1	0.09	0.09	0.08	0.10

In summary, the five variables of x_8, x_{12}, x_{13}, x_{14} and x_{15} have poor correlation with the target variable, so they are excluded. The accuracy and running time before and after the feature selection are calculated; the results are shown in Figure 3.

4. MODEL ESTABLISHMENT

In order to build an SVM model, the sample data is first divided into training set and test set, where the training set is used to determine the Kernel Function and the optimal hyperplane, so as to complete the model building, and the test set is used to test and evaluate the predictive effect of the model. In this process, it is necessary to select the type of Kernel Function and the values of its related parameters, where the linear Kernel Function is the simplest, which only applies to the case where the data can be completely separated, that is, the optimal hyperplane can be found without spatial elevation. The other three Kernel Function are used to deal with the situation that data points are interleaved and cannot be separated, and only they can succeed. After the model is established, the test set is predicted, and its results are compared with the actual situation, so the parameters of the model can be adjusted by feedback to optimize the model. The process steps are shown in Figure 4.

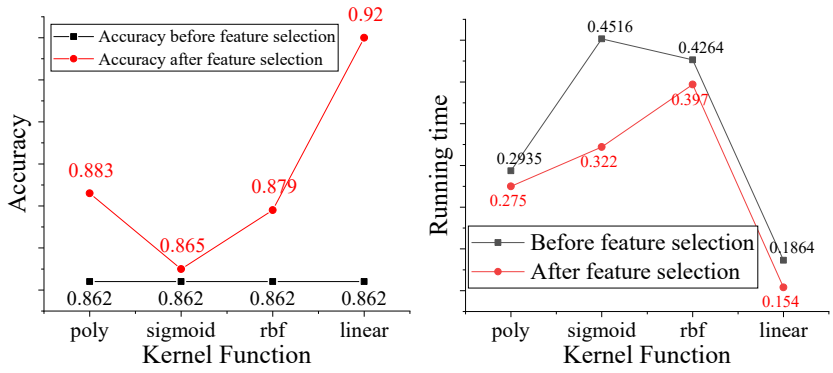

Figure 3. Comparison before and after feature selection.

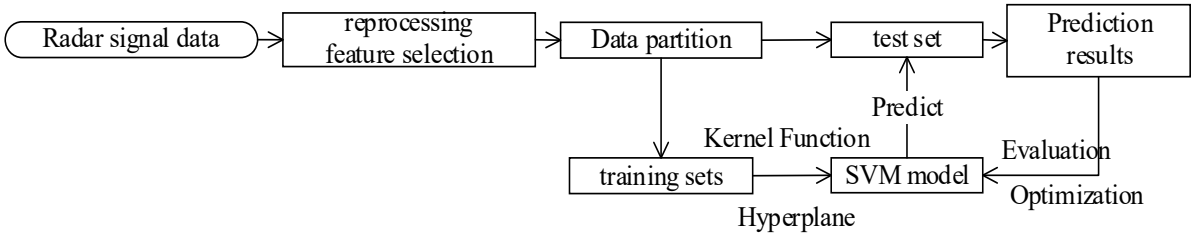

Figure 4. SVM classification process.

4.1 The division of training set and test sets

When dividing the training set and test set, if the proportion of training set is too large, it may lead to the classification being too sensitive to produce a model overfitting, and a small number of test set results are not convincing. On the contrary, if the training set is too small, the data features and information are not fully utilized, resulting in underfitting. In order to determine the optimal partition ratio and ensure the accuracy of the model prediction, the prediction Accuracy when the support vector machine selects four Kernel Function is tested separately under different training set proportion conditions. As shown in Figure 5, experiments have proved that when the training set ratio is equal to 0.8, the Accuracy of the four Kernel Function is at the highest level, so the ratio set to the test set is 8:2.

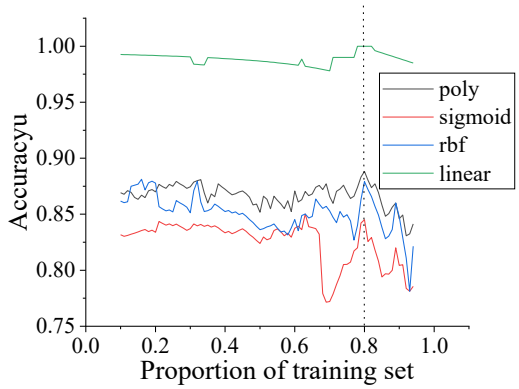

Figure 5. The division of training and test sets.

4.2 The Solution of the Optimal Hyperplane

Since the optimal hyperplane is related to support vectors, to determine the position of the optimal hyperplane, the positions of the support vectors are first determined. Take the two-dimensional space shown in Figure 6 as an example. First, map the data to the two-dimensional space axis to form a point set, and then connect the data points at the outermost periphery of the two categories with a straight line, you will get two polygons, and it can be seen that the end points of the polygon must

contain support vectors. Select one of the endpoints, draw a straight line through it, and make its parallel line tangent to the other polygon, calculating the distance between the two parallel lines. By going through the cycle, it is possible to find a certain point and two straight lines to make the interval largest, and the hyperplane is located in the middle of the two lines.

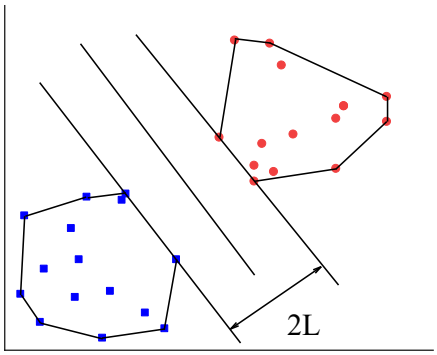

Figure 6. The best way to find a hyperplane.

If the optimal hyperplane cannot be found, the Kernel Function must be used. To prevent the impact of anomalous data points on classification, a small number of data points are typically allowed to cross the baseline, so a certain degree of training error is allowed. In order to measure the tolerance for error, the penalty parameter C is introduced, and the larger C indicates the lower tolerance for error, that is, the narrower the boundary of the hyperplane, so the weaker the generalization ability.

5. MODELS PREDICTION AND EVALUATION

The prediction of the model is actually based on the spatial position relationship between the test set data and the hyperplane equation previously determined to determine the category of the test set data, and the evaluation of the model is to compare the predicted value of the model with the actual value to calculate various evaluation indicators and comprehensively measure the performance of the model. The implementation steps are shown in Figure 7.

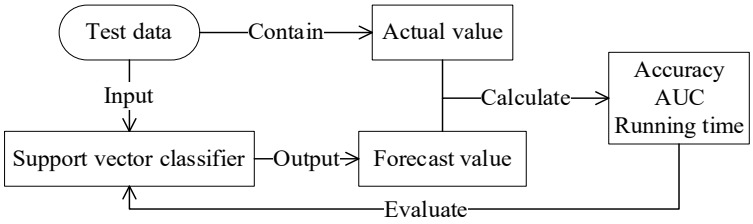

Figure 7. Testing and evaluation process.

5.1 Confusion matrix

A confusion matrix is actually a statistical table of classification results that reflects how well predicted values match actual values, and is often used to measure the accuracy and ability of model classifications[10]. Take the identification of radar countries as an example, there is a binary matrix shown in Table 5.

Table 5. Confusion matrix of two classifications.

Actual category	Forecast category	
A, quantity =T	A: TP	B: FN
B, quantity =F	A: FP	B: TN

Note: T represents the actual number of category A, and TP and FN represent the number of two categories in the prediction results of the model for category A, respectively, and T = TP + FN; similarly, F = FP + TN for category B.

5.2 Evaluation indicators

5.2.1 Accuracy. The prediction accuracy is the most commonly used evaluation index to measure the prediction accuracy of the test set by the model. The calculation equation is

$$Accuracy = (TP + TN)/(TP + FN + FP + TN) \tag{9}$$

Since the training set and the test set are randomly divided according to the scale, so the results obtained are different each time, then there may be differences in the hyperplane determined according to the training set and the prediction results of the test set, which often leads to the accuracy of each prediction always changing. Therefore, the idea of cross-testing is taken here, that is, the original data is divided into 5 parts on average, each time 1 copy is extracted from it as a test set, the remaining four copies are used as training sets, after five different extractions, all the data will have the corresponding prediction value, and the overall prediction accuracy at this time will be used as an evaluation index, which not only makes full use of all the data, but also is more comprehensive and reliable in the evaluation. Calculating the overall accuracy under different Kernel Function conditions separately yields the results shown in Figure 8.

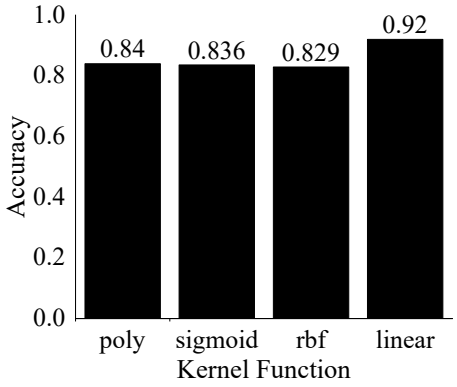

Figure 8. Accuracy comparison of different Kernel Function.

5.2.2 True Positive Rate, False Positive Rate and AUC Value. While accuracy can reflect the predictions made in the overall data, the disadvantage is clear in that it is impossible to know exactly which predictions are wrong or correct, and it is impossible to measure the predictive power of the model for a category. At this time, the True positive rate and False positive rate evaluation indicators are introduced. The TPR calculates the proportion of all that actually belongs to the A category and the model predicts that it is also A, while the FPR calculates the proportion of all that actually belongs to the B category but the model predicts to be A. The calculation equation is

$$TPR = \frac{TP}{T},\ \mathrm{FPR} = \mathrm{FP/F} \tag{10}$$

Obviously, an acceptable model must have the ability to predict both categories. If in a sample consisting of 1 category A and 9 categories B, the model's prediction results are all A, although the accuracy is still as high as 0.90, but the model simply cannot recognize category A, that is, it loses the ability to classify. ROC curve was drawn with FPR as abscissa and TPR as ordinate as shown in Figure 9.

Ideally, high TPR and low FPR are expected, i.e., the closer the ROC curve is to the point (0, 1), the better the performance of the model. To quantify the performance excellence of the model, the evaluation index AUC value is introduced, which is equal to the area below the ROC curve. The result of the integral calculation when selecting different Kernel Function is shown in Figure 10.

5.2.3 Running Time. It is not advisable to blindly pursue the most ideal prediction results and ignore the calculation of time, under the conditions of informationized warfare, whoever can take advantage of efficiency will be able to win the opportunity on the battlefield. Therefore, while improving the performance of the model, we must also pay attention to improving the prediction efficiency of the model. The running time of the four different Kernel Function are shown in Figure 11.

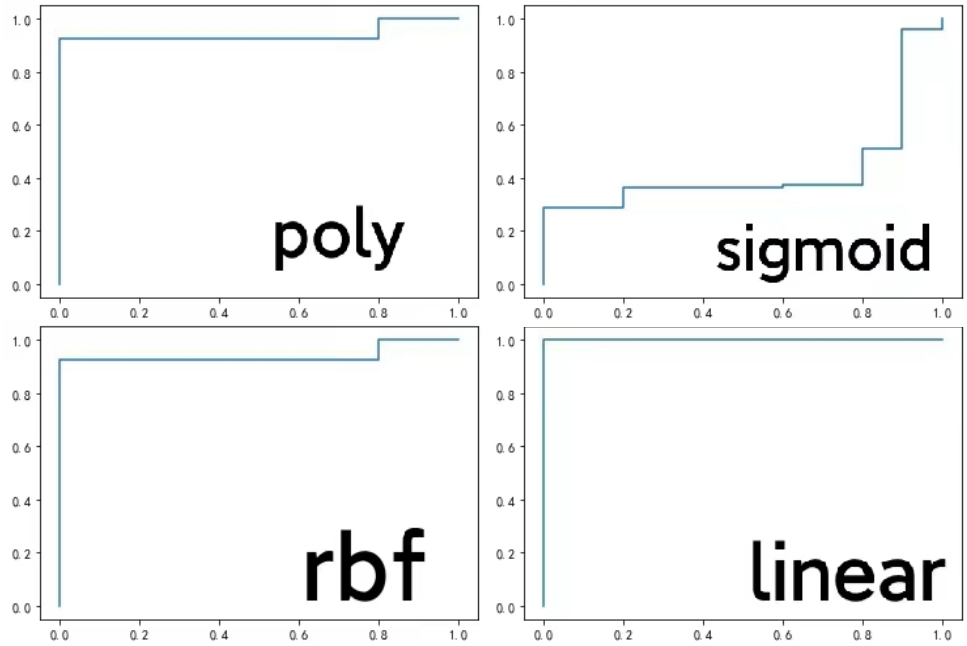

Figure 9. ROC curves for different Kernel Function.

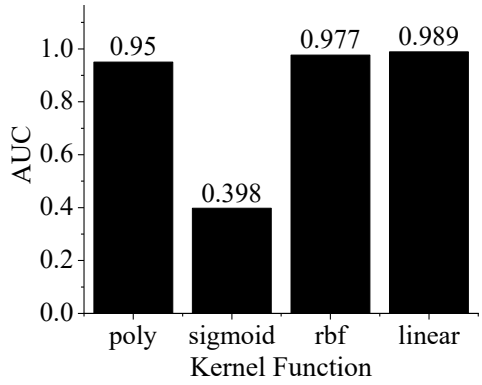

Figure 10. AUC comparison of different Kernel Function.

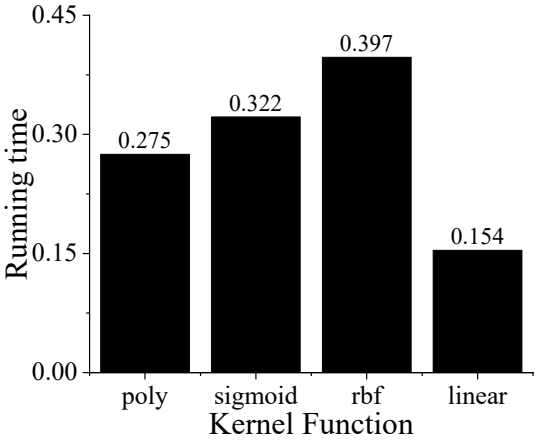

Figure 11. Runtime of different Kernel Function.

6. MODEL OPTIMIZATION

The optimization of the model is actually based on the prediction results to make feedback adjustments, so as to continuously improve the performance of the model, usually reflected in the Accuracy, AUC value increase and the speed of operation.

6.1 Feature importance

From the previous discussion, it can be seen that after removing the five weakly correlated feature variables, both the model performance and the prediction efficiency have been greatly improved. But this may not be enough, because in addition to one target variable, there are still 10 feature variables, which means that the use of support vector machine models to classify radar countries in 10-dimensional space will be implemented, and the complexity of the model is still very high. In order to further reduce the number of feature variables, avoid unnecessary computing resources, and reduce the complexity of the model as much as possible, the importance assessment of 10 feature variables is calculated, and their importance for identifying the country to which the radar belongs is calculated, as shown in Table 6.

Table 6. Importance of characteristic variables.

Variable code	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_9	x_{10}	x_{11}
Degree of importance	0.89	1	1	0.87	1	0.81	1	1	1	0.99

According to the ranking results of the importance degree, it can be seen that the three characteristic variables of x_1, x_4 and x_6 are weaker, they are excluded so that variables reduced to 7 items, and the change of Accuracy, AUC value, and running time of the model are shown in Table 7.

Table 7. Optimizations brought about by feature selection.

Kernel Function	**Accuracy**		**AUC**		**Runtime**	
	Before optimization	After optimization	Before optimization	After optimization	Before optimization	After optimization
Poly	0.84	0.84	0.95	0.975	0.275	0.236
Sigmoid	0.836	0.836	0.398	0.362	0.322	0.287
rbf	0.829	0.829	0.977	0.977	0.397	0.309
Linear	0.92	0.931	0.989	0.993	0.154	0.095

6.2 Parameter tuning

In order to optimize the model to obtain the most ideal classification results, a large number of attempts are made on various Kernel Function and parameters, and the main parameters are shown in Table 8. Here, using exhaustive traversal, the prediction results are tested separately under all possible conditions.

Table 8. Parameter names and meanings.

Parameter name	**C**	**Degree**	**Gramma**	**Coef0**
Meaning	Penalty coefficient	Function dimensions	The parameters of the Kernel Function	Constant terms of Kernel Function

6.2.1 Linear Kernel Function. The parameters of the linear are only the penalty coefficient C, so it is traversed on C to know that when C = 0.1, the prediction Accuracy is 0.931, the AUC value is 0.994, and both evaluation indicators reach the maximum, as shown in Figure 12.

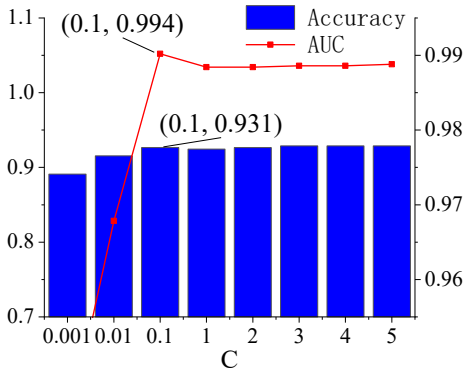

Figure 12. Parameter tuning for linear Kernel Function.

6.2.2 Sigmoid Kernel Function. The parameters of the Sigmoid Kernel Function include the penalty coefficients C, gamma, and coef0. Through exhaustive iteration, it is found that regardless of how the parameters change, the performance of the model has not changed at all, and it is still undesirable.

6.2.3 Gaussian Kernel Function. The main parameters of the Gaussian Kernel Function are C and gamma, and the experiment shows that the optimal result is obtained when C=2, gamma=0.0001, at this time the Accuracy is 0.933 and the AUC value is 0.933, as shown in Figure 13.

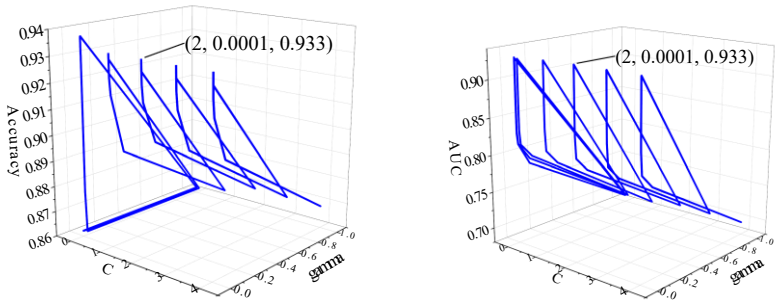

Figure 13. Parameter tuning of rbf Kernel Function.

6.2.4 Polynomial Kernel Function. The parameters of the polynomial Kernel Function are: C, degree, gamma, coef0. It is traversed that the optimal result is obtained when the combination of parameters satisfies the conditions shown in Table 9 with an Accuracy of 0.936 and an AUC value of 0.966.

Table 9. Parameter tuning for poly Kernel Function.

C	Degree	Gamma	Coef0
3	1	0..0001	0

7. CONCLUSION

In this paper, the SVM model is used to identify radar signals and make classified predictions for their countries. The results show that except the sigmoid, the other three Kernel Function can all achieve excellent prediction results, which proves that the SVM has strong classification ability and high prediction accuracy, and only relies on a small number of feature variables to achieve high performance of the model, which can effectively prevent the model from overfitting. The disadvantage is that the calculation process of the support vector machine is complicated, the workload is large, the model complexity is high in the case of large sample data and multi-classification prediction, and a large number of repeated calculations are required when selecting Kernel Function and their parameters.

This paper does not go deep enough in some aspects, such as parameter setting of Kernel Function, multiclass prediction of SVM, etc. Since SVM is a binary classification algorithm, it is necessary to repeat the process of binary classification when dealing with multi-classification problems, and the next step will be to do a detailed study on how to improve the model performance of SVM in the case of multi-classification prediction.

REFERENCES

[1] Zhang, G. W., Luo, D. L. and Li, P., "Research on target classification of battlefield surveillance radar," Fire Control Radar Technology, 47(2), 33-37(2018).

[2] Jin, Q., Wang, H. Y. and Ma, F. F., "Review of classification and identification methods of radar radiation sources," Telecommunications Technology, 59(3), 8-9(2019).

[3] Li, M., Zhu, W. G. and Chen, W. G., "A review on radar radiation source identification based on machine learning," Journal of Ordnance equipment Engineering, 37(9), 2-5(2016).

[4] Chen, B., Liu, J. Q. and Xu, B., "A radar target recognition method based on recurrent neural network model," Journal of Electronics and Information, 38(12), 6-8(2018).

[5] Zhu, J. S., Cong, A. J. and Xu, J., "Research on Radar model recognition method based on decision tree," Electronic World, 1(5), 4-11(2012).

[6] Xu, Y., Gu, Y. and Peng, D. L., "Composite aperture radar image target recognition based on DRGAN and support vector machine," Optics and Precision Engineering, 28(3), 9-16(2020).

[7] Zhao, H. and Li, P. P., "Optimization method for feature selection of machine learning based on proximity gradient," Computer Simulation, 37(11), 3-5(2020).

[8] Song, X. M., Zhang, G. F. and Xing, S. L., "Curriculum correlation analysis method based on data mining," Journal of Shanxi University of Finance and Economics, (2012).

[9] Shen, X. M. and Wang, Y., "Application of mathematical statistics in data analysis," Journal of Harbin Vocational and Technical College, (2019).

[10] Yuan, Z. H. and Yuan, R. Y., "Application of SVM classifier design in radar signal sorting," Electronic Information Countermeasure Technology, 33(6), 5-7 (2018).