# Radar recognition system based on XG-Boost

**Hongxiang Luo, Jianxin Guo[*], Pengxi Fu**

*School of Electronic Information, Xijing University, Xi'an, 710109, China*
*\*guojianxin@xijing.edu.cn*
*[*]Corresponding author*

*Keywords:* Radar recognition, deep learning, XG-Boost algorithm

*Abstract:* With the development of science and technology, radar recognition has gradually entered our lives. With the introduction of deep learning technology in the field of radar recognition, with its strong automatic feature learning ability and end-to-end processing advantages, its recognition accuracy has been further improved. Based on this, a radar recognition system based on XG-Boost is proposed in this paper. The system was used to identify five different materials: air, books, hands, knives and plastic boxes. After a series of experiments, it is found that the recognition accuracy of XG-Boost algorithm is as high as 97.8%, which is higher than the 96.4% of SVC algorithm and 92.8% of GaussianNB algorithm. And the XG-Boost algorithm has achieved 100% recognition rate for air, books, hands and plastic boxes. There was only an error in the identification of the knife.

## 1. Introduction

Radar automatic target recognition technology refers to the technology that uses radar to emit electromagnetic waves to illuminate targets and analyze the obtained echoes to determine the type, model and attributes of the target[1].

In the 1950s, research on radar target recognition had already begun. In the early days, research on radar target characteristic signals mainly focused on the effective scattering cross-section area to evaluate the scattering performance of radar targets. However, since various types of targets have different shapes and properties, it is too simplistic to describe them by just using a unified effective scattering cross-section area, and cannot achieve effective identification of these targets[2]. For decades, with the continuous development of electromagnetic scattering theory and the continuous advancement of radar technology, with the support of advanced modern signal processing technology, many radar target characteristic signals with identification potential have been discovered. This lays the foundation for establishing corresponding target recognition theory and technology.

In recent years, with the introduction of deep learning technology in the field of radar target recognition, this field has achieved rapid development with its powerful automatic feature learning capabilities and end-to-end processing advantages. A typical radar target recognition method is divided into a training stage and a working stage. The training phase is mostly completed offline, and mainly includes data acquisition, preprocessing, feature extraction and classifier training steps. The work phase is performed online, including data acquisition, preprocessing, feature extraction and classification decision-making steps, and finally outputs the category attributes of the target. This article proposes a radar identification system based on XG-Boost, which can better identify and

classify items[3].

## 2. Classification algorithm

## 2.1 XG-Boost algorithm

XG-Boost, proposed by Chen Tianqi in 2016, stands as a gradient boosting tree algorithm. Its foundational principle lies in the minimization of the second-order Taylor expansion of the loss function to determine optimal split points and leaf node outputs in the context of regression trees, thereby refining the model. In tandem, XG-Boost incorporates regularization terms, including the count of subtrees and values assigned to leaf nodes, within the loss function, presenting a holistic consideration of model complexity[4]. Regarding computational efficiency, XG-Boost markedly enhances its modeling efficiency through distinctive strategies such as approximating regression tree bifurcation points, parallelizing sub-node computations, and leveraging second-order convergence. These attributes collectively contribute to its superior efficiency relative to conventional Gradient Boosting Decision Trees (GBDT) methodologies. The objective function expression for this model is[5]:

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{K=1}^{K} \Omega(f_k)$$

(1)

The objective function consists of two parts, the first part is the loss function, and the second part is the regularization term. The regularization term is obtained by adding the regularization terms of K-trees.

XG-Boost model optimization does not directly optimize the entire objective function, but optimizes the objective function step by step, first optimizing the first tree, then optimizing the second tree, and until the last tree[6]. The available mathematical formula is expressed as:

$$\hat{y}_i^{(0)} = 0$$

(2)

$$\hat{y}_i^{(1)} = \hat{y}_i^{(0)} + f_1(x_i)$$

(3)

$$\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i)$$

(4)

$$\hat{y}_i^{(t)} = \sum_{K=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

(5)

For the regularization of a tree, it can be expressed as a mathematical formula:

$$f_t(x) = \omega_{q(x)}, \omega \in R^T, q; R^d \to \{1, 2, \cdots, T\}$$

(6)

For a tree with T leaf nodes, the values of these T leaf nodes form a T-dimensional vector w. Among them, the mapping function q(x) maps the sample to a leaf node from 1 to T, which actually represents the structure of the CART tree. Therefore, ωq(x) naturally represents the predicted value of the tree for sample x. Then the regularization term of XG-BOOST is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} W_j^2$$

(7)

The parameters $\gamma$ and $\lambda$ can be adjusted and configured based on the specific situation. A larger value of $\gamma$ indicates a preference for obtaining trees with simpler structures, implying a higher penalty for trees with a greater number of leaf nodes. Similarly, a larger value of $\lambda$ also signifies a preference for simpler tree structures[7].

## 2.2 Support Vector Classification (SVC) algorithm

SVM is a two-category classifier that only supports two categories of classification. However, in some cases, multiple SVM can be connected in series to achieve multi-classification purposes[8]. SVC is a special case of SVM, which is mainly used to solve classification problems. In classification problems, the goal of SVC is to find a hyperplane that splits the data into two categories, and this hyperplane maximizes the distance between the two categories. At the same time, the conditions for correct classification are ensured. In practical applications, model performance can be optimized by using different kernel functions, adjusting regularization parameters, etc. SVC performs well in processing high-dimensional data and complex decision boundary problems, but it may require more computing resources for large-scale data sets and high-dimensional features.

For linearly separable cases, the classification hyperplane can be defined as[9]:

$$f(x) = W \bullet X + b = \sum_{j=1}^{d} \omega_j x_j + b \tag{8}$$

If the input sample value x belongs to the negative class, $f(x) < 0$; otherwise, when the input sample value x belongs to the positive class,, then $f(x) \geq 0$[10].

$$y_i f(x_i) = y_i(W \bullet x_i + b) \tag{9}$$

In the above formula, W represents the weight coefficient vector of the classification surface; b is the domain value of the classification. W·X represents the inner product of W and /‖W‖, and the sum of the minimum distances of the two types of samples to the hyperplane 2/‖W‖ is the largest. Therefore, the optimal hyperplane should satisfy the following constraints:

$$y_i(W \bullet x_i + b) - 1 \geq 0, i = 1, \cdots, n \tag{10}$$

Next, assuming that a* is the optimal solution of a, and then w* and b* are the optimal solutions of W and b respectively, the following system of equations can be listed：

$$\begin{cases} \omega^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i = \sum_{\sup portvectors} \alpha_i^* y_i x \\ b^* = 1 - \omega^* \bullet x_i \end{cases} \tag{11}$$

In the above system of equations, w* represents the weight, ai*represents the optimal Lagrangian coefficient, and b* is the optimal threshold. Solving the system of equations, the optimal classification decision function can be obtained:

$$f(x) = \text{sgn}\left[ \sum_{j=1}^{n} \alpha_i^* y_j x_j \bullet x + b \right] \tag{12}$$

The function sgn( · ), known as the sign function, returns 1 for positive numbers, -1 for negative numbers, and 0 for zero. It is employed to determine the classification of samples, assigning them to different categories based on their sign[11].

## 2.3 GaussianNB algorithm

In this paper, GaussianNB is used as the comparison algorithm. GaussianNB is a variant of the Naive Bayes algorithm, specifically suitable for situations where the features are continuous variables and conform to the Gaussian distribution. In this algorithm, each feature is assumed to be normally distributed under a given category. Classification predictions are made by calculating the mean and variance of each feature in each category, i.e., the parameters of the model. The parameter estimation

stage involves calculating the mean and variance of each feature, while in the classification prediction stage, the posterior probability of each category is calculated through the Gaussian probability density function, thereby selecting the category with the largest posterior probability as the classification result of the sample. The algorithm flow is shown in Figure 1.GaussianNB performs well in handling classification problems of continuous features, and is especially suitable for scenarios where data obeys a normal distribution.
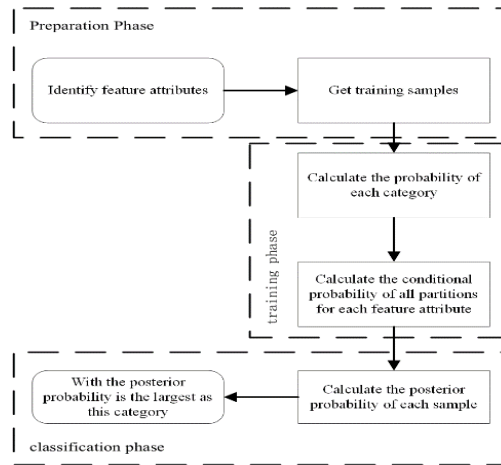


Figure 1: GaussianNB algorithm flow chart

## 3. Experimental process and results

In order to ensure the accuracy of the experimental results, we need to perceive the object at a certain distance and incident angle. Since the reflection and emission characteristics of different materials will affect the intensity of the received radar signal, there are multiple overlapping reflected signals at the same time, thereby increasing the discernibility and reliability of the signal. This article uses five objects of different materials for experiments: air, books, hands, knives and plastic boxes. The radar is used to transmit and receive these objects, and then the received signals are processed, and the curves of the maximum cross-section, minimum cross-section and average cross-section of each object are obtained, as shown in Figure 2.
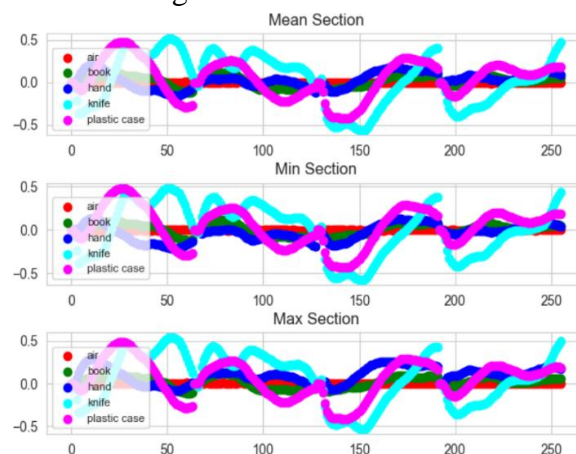


Figure 2: Features of each signal after normalization

After that, we conduct a series of experiments on the data processed as above. XG-Boost was compared with the other two algorithms and found to have higher recognition accuracy. The overall recognition accuracy is shown in table 1. It can be seen from the table that the recognition accuracy

of XG-Boost algorithm is higher than SVC algorithm and GaussianNB algorithm. At the same time, in order to display the experimental results more intuitively, a confusion matrix is introduced to show its recognition accuracy. The abscissa and ordinate respectively represent the five objects to be classified: air, books, hands, knives, and plastic boxes. The data in the matrix represents the recognition accuracy, and the darker the color in the diagonal grid, the higher the recognition accuracy.

Table 1: Accuracy of each classifier

|  | Accuracy | Error |
|---|---|---|
| XG-Boost | 97.8% | 2.3% |
| Support Vector Classification | 96.4% | 3.6% |
| GaussianNB | 92.8% | 7.2% |

Figure 3 shows the Gaussian NB classification results. Although its comprehensive accuracy reaches 92.8%. However, it is not difficult to see from the confusion matrix that the accuracy of this algorithm when identifying books is low, only 75%. There are also certain errors on knives with metal materials, and the recognition accuracy is only 89%.
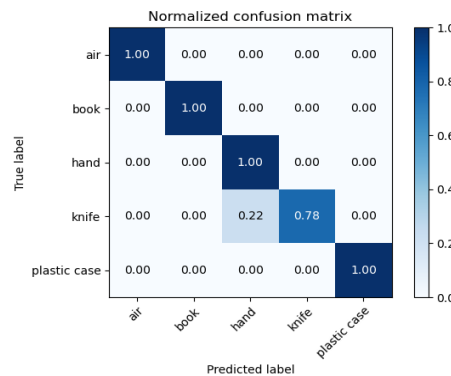


Figure 3: Results of GaussianNB algorithm

Next, the SVC algorithm is used to classify it. The confusion matrix of the classification result is shown in Figure 4, and its comprehensive accuracy also rises to 96.4%. And compared with the Gaussian NB algorithm, its recognition rate for books reaches 100%. However, its knife recognition rate is only 78%, which is far lower than Gaussian NB's 89%.
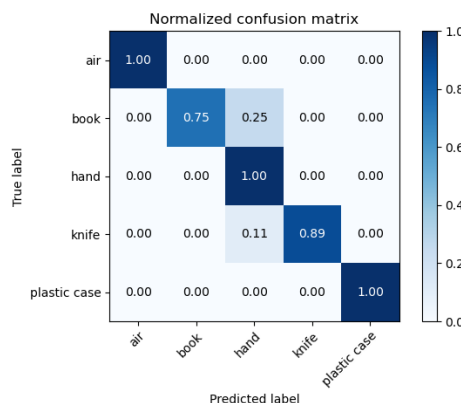


Figure 4: Results of SVC algorithm

Finally, the XG-Boost algorithm proposed in this paper was tested, and the confusion matrix for its classification results is shown in Figure 5. The overall accuracy reached 97.8%. Moreover, the recognition accuracy for four objects—air, books, hands, and plastic boxes—reached 100%. There

were some errors in the identification of knives, with an accuracy rate of 89%. Thus, it can be inferred that the XG-Boost algorithm significantly improves the accuracy of radar recognition.
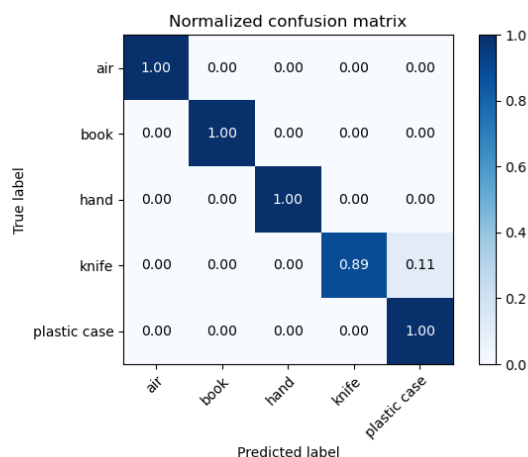


Figure 5: Results of XG-Boost algorithm

## 4. Conclusions

In this paper, we introduce some deep learning algorithms to improve the accuracy of radar recognition and classification, and verify them through a series of experiments. The experiment proves that the combination of deep learning and radar recognition classification is a stable and reliable classification method, in which XG-Boost algorithm shows obvious advantages. With the further development of technology, radar identification and classification systems will be more and more widely used in daily life, such as transportation, industrial manufacturing, medical treatment, meteorology and so on. However, there are also some areas that need to be improved, such as when identifying some items with metal, its recognition accuracy will decline, and it can not reach 100% recognition accuracy. Therefore, in the future, we can consider using the fusion of multiple algorithms to further improve the accuracy.

## Acknowledgements

## References

[1] Dandan Gu, Yi Liao, Xiaobing Wang. Progress and Thinking of Intelligent Recognition Technology Guided by Radar Target Characteristics Knowledge[J]. Guidance & Fuze, 2022, 43 (04): 57-64.
[2] Marlin. Survey of Radar Target Recognition Technology[J]. Modern Radar, 2011, 33(06): 7.
[3] Pengcheng Guo, Jingjing Wang, Longshun Yang. Current status and prospect of radar ground target recognition technology [J].Aero Ordnance, 2022, 29(02):1-12.
[4] Asif N , Tariq A , Ghulam M , et al.A novel technique for detecting electricity theft in secure smart grids using CNN and XG-boost[J].Intelligent Systems with Applications, 2023, 17.
[5] Lina YANG, Kaixue YAO, Yong HE et al. Research on road condition sensing method based on SmoteEnn_XGBoost model [J]. Intelligent Computer and Application, 2021, 11(11):137-142+147.
[6] Miaomiao Li. Drug relocation prediction based on XG-B00ST and multiple data sources [J]. Software Guide, 2020, 19(02):110-113.
[7] Kangmo Jung. Support Vector Machines for Unbalanced Multicategory Classification [J].Mathematical Problems In Engineering, 2015, 1(1).

[8] Yingran S , Chandra G , Yang Y , et al.Recurrence prediction of lung adenocarcinoma using an immune gene expression and clinical data trained and validated support vector machine classifier.[J].Translational lung cancer research, 2023, 12(10):2055-2067.

[9] Yitou Li. Improved classification algorithm based on support vector machine[J].Computer System Applications, 2019, 28(10):145-151. DOI:10.15888/j.cnki.csa.007080

[10] Danling Chen. Research on fault diagnosis method of rotating machinery based on wavelet packet and support vector machine [D]. Jiangxi University of Science and Technology, 2010.

[11] Xin LIU, Haochen WANG, Yuhu HUANG. Recognition of telecom fraud information based on plain Bayesian classification [J]. Computer Age, 2023, (04):29-32+38. DOI: 10.16644/j.cnki.cn33-1094/tp.2023.04.006