A method for feature extraction based on SVD and machine learning

Wei Mao, Shuxian Huang, Xin Liu, Hongyan Liu, Jiaqi Liu, Yi Shu

1National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics, 100076, China
2Beijing Institute of Precision Mechatronics and Controls, 100076, China
3Beijing Institute of Long March Vehicle, 100076, China
mwmaomw@139.com

Abstract. By studying the shortcomings of feature, which extracted from Radar-Cross Section (RCS) using mathematical and statistical method, using the idea of extracting abstract features in image recognition and speech recognition by artificial intelligence for reference[2][3]. This paper explores the possibility of extracting abstract features of target’s RCS sequence, and proposes an abstract feature extraction method of RCS sequence based on singular value decomposition (SVD) feature decomposition. Because of the poor interpretability of abstract features, four different machine learning algorithms are used to classify the extracted abstract features. The experimental results show that the machine learning algorithm can classify different types of spatial objects with high accuracy, which shows that the RCS features of different spatial objects can be characterized by abstract features.

1. Introduction
Space target recognition is getting more and more attention by researchers, with the human exploration of outer space development in-depth. The main task of space target recognition is the accurate detection and recognition of space target, to determine the shape, size and orbit parameters and other important information, and to classify and recognize the target. Space target recognition is worth studying, and it can not only identify the type of space target, warn and prevent possible collisions and intrusions, but also improve the ability of outer space exploration.

Feature extraction is very important, which directly affects the accuracy of target classification and recognition. The RCS reaction of the target scattering ability of electromagnetic wave, is an important aspect of radar target characteristic[1][4]. However, RCS is extremely sensitive to the attitude of the target, it is almost impossible to identify the target by using RCS data directly. The mathematical and statistical characteristics of RCS are often used as the basis of target recognition, in the traditional research of target recognition based on RCS. This paper introduces the shortcomings of mathematical and statistical characteristic in target recognition, and proposes an abstract feature extraction method of RCS based on machine learning.

2. The Deficiency of Mathematical and Statistical Feature on RCS
The mean value of the RCS sequence can reflect the electromagnetic scattering ability of the target, and the variance can reflect the attitude stability of the target. Based on this, the target type can be preliminarily determined. However, in some cases, it is difficult to recognize the target through the
mathematical and statistical feature of RCS sequence. In this section, we simulate a typical scene to obtain the dynamic RCS of the target and extract its mathematical and statistical features, and analyze the shortcomings of mathematical and statistical features in target recognition.

2.1. simulation experiment
In the scene shown in Figure 1, STK software is used to generate a typical space target scene, which is detected by two radars deployed at the under star point of the space target trajectory. Due to the different radar positions, the incident angles of the two radars for detecting space targets are also different. As a result, the same space target’s RCSs, which are detected in the different two radars, are different, as shown in Figure 2. The statistical characteristics of the RCSs detected by the two radars are showed in the Figure 3.

Fig. 1 Detection scene of space target by radar in different deployment positions
Fig. 2 Target RCS sequence
Fig. 3 PDF and CDF of space target
RCS of radar target is very sensitive to the attitude of the target, so RCS is a fluctuating variable. The RCS fluctuation model can be expressed as the probability density function (PDF) of RCS in a limited attitude angle range\(^5\). Cumulative probability density function (CDF) of target’s RCS can be obtained by PDF. Therefore, we can compare the CDFs of two RCS sequences and determine whether the CDFs of two RCS sequences obey the same distribution by non-parametric hypothesis test. The commonly used test method is Smirnov test method\(^6\).

Let’s assume \((X_1, X_2, \ldots, X_n)^T\) as the sample from a population \(X\) with continuous distribution function \(F(x)\), \((Y_1, Y_2, \ldots, Y_n)^T\) as the sample from a population \(Y\) with continuous distribution function \(G(x)\), and the two samples are assumed to be independent of each other. To test the hypothesis:

\[H_0 : F(x) = G(x) \leftrightarrow H_1 : F(x) \neq G(x), -\infty < x < \infty\] (1)

Let \(F_n(x)\) and \(G_n(x)\) be the empirical distribution functions corresponding to the two samples, respectively, and make statistics:

\[D_{n_1,n_2} = \sup_{-\infty < x < \infty} \left| F_n(x) - G_n(x) \right|\] (2)

For a given significant level \(\alpha\), make \(n = \frac{n_1n_2}{n_1 + n_2}\), \(D_{n,\alpha} = \frac{\lambda_{n,\alpha}}{\sqrt{n}}\), if \(\hat{D}_{n_1,n_2} > D_{n,\alpha}\), then reject hypothesis \(H_0\). If \(\hat{D}_{n_1,n_2} \leq D_{n,\alpha}\), then accept the original hypothesis \(H_0\), at the significant level \(\alpha = 0.05, \lambda_{n,\alpha} = 1.36\).

Through this hypothesis test, we can verify whether the PDFs of two RCS curves come from the same distribution. If they do not come from the same PDF, we can conclude that the two RCS sequences do not come from the same target.

Under the condition that confidence is 95%, \(D_{n,\alpha} = 0.014\), \(\hat{D}_{n_1,n_2} = 0.27\) was calculated by using Smirnov test. Namely \(\hat{D}_{n_1,n_2} \leq D_{n,\alpha}\), then \(H_0\) is rejected, that is to say, the CDFs of the same space target’s RCS, which detected by two radars at different locations, do not come from the same distribution.

Calculate the means and variances of RCSs, which detected by the two radars, can get the results shown in table 1.

| Radar1 detected RCS | Radar2 detected RCS |
|---------------------|---------------------|
| Mean(dB)            | Variance            |
| 25.51               | 12.53               |
| 28.53               | 9.45                |

2.2. Brief Summary

Through the simulation analysis in this section, it can be found that the RCS of space target is extremely sensitive to its attitude, so the RCS measurement results of radar deployed in different positions are different for the same space target. Through the analysis of the traditional mathematical and statistical characteristics, it is found that the mean value of RCS measured by two radars differs by 3dB, and the CDF does not subject to the same distribution, that is to say, there is not a RCS statistical feature that can be used to uniquely represent a spatial target.

3. Feature Extraction Method of RCS Sequence Based on Artificial Intelligence

3.1. Brief introduction of feature extraction using SVD
In recent years, with the development of artificial intelligence technology, such as deep convolution neural network, it has become more and more practical to extract abstract features of data by transformation method, and has achieved good application results in image recognition, speech recognition and other fields. Starting from this idea, we explore the possibility of using abstract features to represent spatial objects. In this paper, we use SVD algorithm to extract the abstract features of RCS sequences of spatial objects\cite{7}. SVD algorithm can extract the energy features of RCS sequences very well. Large samples are generated from a large number of RCS simulation data, and the abstract features of samples are extracted by SVD. Because of the poor interpretability of these abstract features, machine learning algorithm is used to classify these features, and the classification accuracy is used to represent the quality of feature classification. In order to prevent the accidental difference of classification results caused by a single machine learning algorithm, four commonly used machine learning algorithms are used to classify the abstract features of RCS, and the results indirectly reflect the quality of abstract features extraction.

SVD is the most common matrix decomposition technology. Matrix decomposition can represent the original matrix in a new and easy-to-process form. In many cases, a small segment of data carries most of the information in the data set. Other segment is either noise or irrelevant information. SVD can approximate the original data and extract important features from it. By retaining 80% to 90% of the energy of the original data matrix, it can remove noise and retain important features.

The common machine learning methods are used to classify and recognize RCS sequences. This paper mainly adopts four different machine learning methods: Naive Bayesian method, k-nearest algorithm, support vector machine(SVM) and random forest\cite{8}\cite{9}.

3.2. Experimental results and analysis
In this experiment, the shape of typical space objects, including cone, column and square column, is selected as shown in (a) in Fig. 4, and three space objects of different sizes are selected as shown in (b) in Fig. 4. Assuming that the surface materials of these space objects are the same, the RCS static characteristics of the target at 10 GHz frequency are shown in Fig. 5.

![Fig.4 Size and shape of the space target](image)

(a) Targets of the same length with different shapes (b) Targets of the same shape with different lengths

Radars are deployed at different locations below the orbit of spacecraft to detect targets. Each RCS sequence detected by radar is taken as a sample, and the RCS samples detected by different radars are combined into a matrix, the RCS data with insufficient length are zero-padding. In this way, the abstract features of these targets can be extracted by SVD method, and the corresponding abstract features of each RCS sequence can be obtained. At the same time, these abstract features can be labelled to represent the corresponding target type of RCS. The abstract features and corresponding labels are divided into two parts, one is training data, the other is test data, in which the ratio of training data and test data is 7:3. The training data are used to train different machine learning algorithms, and the model after training is used to test the test data. The target recognition probability under this feature can be obtained accurately. Among them, the highest recognition probability of machine learning can indicate the quality of abstract feature extraction. When the abstract feature extraction is reasonable, machine learning algorithm can distinguish the target of the feature with high recognition probability. Three different scenarios are simulated and analyzed as following.
3.2.1. Scenario 1

In the scenario of Fig. 1, the radar is deployed in the range of 3000km on both sides of the space target orbit, and the RCS of the cone targets with different lengths is simulated on the same orbit. The simulation data are used to extract the abstract features according to the previous process. Then the accuracy of the machine learning algorithm is used to verify the abstract features. Since the abstract feature is a multi-dimensional feature, the first several-dimensional feature occupies a larger weight in the energy of the whole signal, and the joint distribution of the first and second dimension features are displayed, as shown in Figure 6. For cone targets of different sizes, the classification results obtained by Naive Bayesian, KNN, SVM and Random Forest algorithms are shown in Figure 7.

As can be seen from the results, the maximum probability of accurate recognition is 75%. In Figure 6, the horizontal axis represents the first feature and the vertical axis represents the second feature. The joint distribution of the first and second dimension features of the three cone targets is not particularly different. It can be seen from the static RCS of the target that the RCS of the three targets in the range of $-70^\circ \sim 70^\circ$ has little difference, which leads to the little difference of abstract features extracted. However, from the 75% accurate recognition rate, we can see that abstract features still have a certain degree of discrimination.
3.2.2. **Scenario 2**
The RCS of cone targets with different lengths, which in different orbits, are simulated. And the radars’ deployment are the same of the scenario1, but the apogees of the orbits are different. The abstract features of these simulation data are extracted according to the previous process. Then the accuracy of the machine learning algorithm is used to verify the abstract features. The joint distribution of the first and second dimension features are displayed, as shown in Figure 8. The classification results obtained by Naive Bayesian, KNN, SVM and Random Forest algorithms are shown in Figure 9.

From the recognition rate in Figure 9, it can be seen that the maximum probability of accurate recognition is 72.9%. Compared with the previous scenario, the detection angle span of different radars for the same target in different orbits is larger. Under the same feature extraction method, the extracted feature is slightly worse, which leads to a decrease in the probability of accurate recognition.

![Fig.8 The joint distribution of the first and second dimension](image1)

![Fig.9 Accurate recognition rate of machine learning algorithms](image2)

3.2.3. **Scenario 3**
The RCS of space targets with different shapes, which in different orbits, are simulated. And the radars’ deployment and the orbits of the space targets are the same of scenario 2. These simulation data are used to extract abstract features according to the previous process, and then the accuracy of the machine learning algorithm is used to verify the quality of abstract feature.

As can be seen from the results, the maximum probability of accurate recognition is 96%. From the joint distribution of the first and second dimension features in Fig. 10, it can be seen that the differences of the joint distribution of the three targets is large, so the probability of accurate recognition is greatly improved.

![Fig.10 The joint distribution of the first and second dimension](image3)

![Fig.11 Accurate recognition rate of machine learning algorithms](image4)
3.3. Brief Summary
From the simulation results in this section, it can be seen that different machine learning algorithms are used to classify the RCS, and the classification results are good. This shows that the target feature extraction method based on SVD can better reflect the similarity of the same target under different radar detection conditions, at the same time, it can also better reflect the difference of characteristics between different targets.

4. Conclusion
Starting from the feature extraction of RCS for space targets, this paper uses the idea of abstract features extraction for image and speech in artificial intelligence for reference, and puts forward the idea of abstract features extraction for target RCS as the basis of recognition. The energy feature in RCS is extracted by SVD algorithm, and the abstract feature is used as the basis of target recognition. Different machine learning algorithms are used to classify different features. The highest accurate recognition probability is 96%. This shows the feasibility of extracting the abstract feature of RCS based on SVD algorithm.

Because the interpretability of abstract features is poor and there are many methods to extract abstract features of objects, the feasibility of extracting RCS abstract features of objects through different algorithms such as convolution network will be studied in the future, and the interpretability of abstract features of objects will be studied to explore an abstract feature extraction method with better interpretability and better classification effect.

References
[1] E.F.Knott et al.. Radar cross section, 2nd edition. Dedham[M], MA: Artech House, 1993.
[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition. In CVPR 2016.
[3] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. In ECCV, 2014.
[4] Peikang Huang. Radar target characteristics[M]. Electronic Industry Press.
[5] Xiaojian Xu. Radar system and its information processing[M]. Publishing house of electronics industry.
[6] Yiming Shi. Mathematical statistics[M]. Science Press.
[7] A New Approach for Face Recognition Based on Singular Value Features and Neural Networks[J]. 2004, 32(1): 171-173 (in Chinese).
[8] Wu X, Kumar V, Ross Quinlan J, et al. Top 10 algorithms in data mining[J]. Knowledge and Information Systems.
[9] Wolpert D H. The Supervised Learning No-Free-Lunch Theorems[M]. Springer London.