

# The simulation of the optimal classification mathematical model based on chaos probabilistic analysis

Chunyan Lv, Chuan He, Erwei Wang

Zaozhuang Vocational College of Science&Technology, shandong,tengzhou,277500

**Keywords:** chaos; probability analysis; classification; mathematical model

**Abstract.** This paper proposes the establishment and simulation of one optimal classification mathematical model based on chaos probability analysis. A kind of recurrence plot chaotic analysis algorithm is applied to recognize the mathematical classification features and the improved neural network algorithm is used to design the mathematical classification model. The simulation experiments illustrate the proposed classification mathematical model can effectively classify and reorganize each types of data with satisfactory performance. The error rates are reduced averagely by 13% compared with the neural network classifiers, which reflects the superior performance of the algorithm. The proposed algorithm can be effectively applied in the classification reorganization fields, such as data mining, troubleshooting diagnosis and targets identification.

## 1 Introduction

With the growing development of the data processing and information processing, massive amounts of data information should be effectively classified and reasonably regulated in order to achieve the effective management of the data information. Especially the development of the network technology and large amounts of data need effective mathematical models and algorithms to optimize the data classification. The optimal classification of the data focuses on original data. The similar characteristic data will converge to the output of the classifier with some algorithms and manners. The classification algorithms can reveal the essences of the original data to classify the data, which can be effectively applied in the military and civilian fields, such as engineering troubleshooting, radar target identification, classification networking technology and oil and water two-phase flow diversion. They have wide applications in prospects<sup>[1-3]</sup>.

## 2 The features recognition with chaos analysis in the recurrence plots

The recurrence plots chaotic algorithm uses the phase space reconstruction of data non-linear time sequences as the foundation. Meanwhile, the phase space reconstruction is the key and guide for the non-linear time sequences which can make the scale time sequence can be expressed in the multi-dimension auxiliary space. The core of the phase space reconstruction is to find the proper embedded dimension  $m$  and time delay  $\tau$ . This paper uses the fake nearest neighbor method and mutual information algorithm to select the proper embedded dimension  $m$  and time delay  $\tau$ . The obtained information confirmation system is

$$\mathbf{x}_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}) \quad i = 1, 2, \dots, N - (m-1)\tau \quad (1)$$

where  $\{\mathbf{x}_n\}_{n=1}^N$  is the database information flow time sequence with the length of  $N$ .

The recurrence plot (RP) is an important method to analyze the periods, chaos and non-stationary of the time sequences. The embedded vector  $\mathbf{x}_i$  in the phase space reconstruction can reveal the changing rules of the data with graphical manners. The recurrence plots can be used to test the stability of the database information flows and reveal the interior structure of the time sequences to obtain the relevant similarities, information volumes and the predictive priori knowledge. The analysis of the recurrence plots is intuitive and qualitative which can directly observe the time sequence database information flows are random disordered structure with recognizable modes. If the database information flows are random, the ransom classification methods should be considered to recognize the fixed patterns in order to achieve the performance of chaotic analysis and abnormal

detection. If the database information flows have deterministic components, the distribution of the dotted lines in the RP will have some rules, such as there are regular lines around the diagonal lines, which illustrates the database information flows have some identification and predication capabilities. They can be used for the non-linear quantitative research and analysis to search the effective characteristic values, which provide the foundation for the further classification.

The detailed procedures of the recurrence plots algorithm are as follows.

$R(i, j)$  is used as the horizontal axis and  $j$  is used as the vertical axis to plot the coordinate image  $R(i, j)$  in which  $i, j$  are referred as the time sequence  $i$  and  $j$  respectively. The obtained image is recurrence plot wherein the recursive values  $R(i, j)$  is expressed by

$$R(i, j) = H(\varepsilon_i - d_{ij}), \quad i, j = 1, 2, \dots, N \quad (2)$$

where  $\varepsilon_i$  is cutoff distance which can be fixed values or changing with  $i$  to make the balls with the radius of  $\varepsilon_i$  contain some neighbor numbers. In the equation,  $H(\cdot)$  represents the Heavside functions.  $d_{ij}$  is the distance between the  $x_i$  of  $i$ -th point and  $x_j$  of  $j$ -th point in the reconstructed phase space.

$$d_{ij} = \|x_i - x_j\| \quad (3)$$

where  $i = 1, 2, \dots, N - (m-1)\tau$ ,  $j = 1, 2, \dots, N - (m-1)\tau$ .

According to equation (2),  $R(i, j)$  is either 0 or 1. The recurrence plots are essentially the time sequences described by the black/white dotted lines. The corresponding structures of the recurrence plots describe the complexity and regularity of the signals.

### 3 Design of the classification model

(1) Some specific data which need be classified are used to design the specific features index system which can be used to collect the data.

(2) The abnormal data are eliminated from the collected data and the mean method is used to substitute for data pre-processing.

(3) The pre-processed data are used to reconstruct the phase space. The delay time ( $\tau$ ) and embedded dimension ( $m$ ) of the data sequences are determined which are used to reconstruct the time sequence samples in order to recover the chaotic data to regular sequence data.

(4) The reconstructed chaotic data are divided into two parts-training sample and test sample in which the training sample is used to establish the classification model and test sample is used to test the established model.

(5) The training samples are input into the BP neural network for learning and the genetic algorithm is used to optimize the parameter connection weights ( $w$ ) and threshold ( $\theta$ ) in the BP neural network in order to establish the optimal non-linear classification model.

(6) The established optimal non-linear classification model is used to classify the sequence test samples to verify the model.

### 4 Simulation experiments and analysis of the optimal classification performance

In order to testify the classification performance of the mathematical model, the library samples originally collected from the noise fields are used as experimental data for mathematical modeling and simulation. The sample data libraries are from four types of mechanical noise radiated noise fields respectively which are defined as database1, database2, database3 and database4 accordingly as shown in Table 1.

**Table 1 The experimental data information in the experimental databases**

Databases	No. of categories	Scale of the data	Dimensions
Database1	3	1500	2
Database2	4	2000	2
Database3	5	2200	2
Database4	6	2500	2

The classification error rates in the data classification are projected to a set of probability density function by the channel mapping function approaches of the data classifiers. The probability density function can assign the utility probability for each classification frequency. The specific random number generation sequences with significant essential features are generated by chaotic probability analysis which can meet the previous random FM requirements of the probability density in order to optimize the data classification and mathematical model construction. The traditional neural network classifier and fuzzy c-mean (FCM) data clustering algorithm are used to classify 4 sets of experimental data and the error rates are used as the evaluation index system. The average error rates of the three algorithms with 50 times of experiments are as shown in Table 2. In the table, the error rates by the proposed chaotic probability analysis optimization classification method are less than those in neural network classification and fuzzy c-mean data classification in which it is reduced by 13% compared with neural network classifier and 6% compared with fuzzy c-mean classifier reflecting the superior performance of the proposed algorithm.

**Table 2 The index evaluation of the average error recognition rates with 50 times of experiments by the 3 algorithms**

Databases	Neural network Classification	FCM Classification	Chaos Probability Analysis
Database1	0.03424	0.03246	0.04021
Database2	0.54352	0.53264	0.63278
Database3	0.97547	0.695547	1.086325
Database4	0.62054	0.537547	0.76274

## 5 Conclusions

Based on the chaotic probability analysis method, this paper proposes the establishment and simulation of an optimal classification mathematical model which uses a kind of recurrence plot chaotic analysis algorithm to recognize the mathematical features. The improved neural network algorithm is used to design the mathematical classification model. The chaotic probability analysis can generate the specific random number sequences with significant essential characteristics which meets the random FM requirements with previous probability density. They can optimize the data classification and establish the mathematical models. The real datasets originally collected from the noise fields are used to build the databases as the experimental data for mathematical modeling and simulation. The sample databases are from 4 types of mechanical noise radiated noise fields. The simulation experiments illustrate the proposed classification mathematical model can effectively classify and reorganize each type of data with satisfactory performance. The error rates are obviously reduced compared with traditional neural network classifier and fuzzy c-mean data classifier in which it is reduced by 13% compared with neural network classifier and 6% compared with fuzzy c-mean classifier. It reflects the superior performance of the algorithm which can be effectively applied in classification identification fields, such as data mining, troubleshooting diagnosis and target identification.

## References

- [1] Wang Zhongcai Li Yongbi. Intrusion detection system based on data mining research [J]. Science and technology, 2012, 28 (8) : 150-152.

- [2] Song Shucui. Facing Web data mining technology in the research and application of personalized recommendation methods of website optimization [J]. Science and technology, 2012, 2 (28) : 118-119.
- [3] Hu Guangbo, zhou yong, nonlinear and uncertainty of ship radiated noise test [J]. Journal of ship electrical engineering, 2010, 30 (10) : 150-153.
- [4] Dunn J C A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well - Separated Clusters. Q.c ybernet, 1974 (3) : 32-571.
- [5] Li Chaoshun jian-zhong zhou, remained, etc. Fuzzy clustering analysis method based on chaos optimization [J]. Journal of system simulation, 2009.21 (10) : 2977-2980.