

Design and Implementation of Fishery Forecasting System Based on Radial Basis Function Neural Network

Yuan Hongchun¹, Wang Jintao¹, Chen Ying², Chen Xinjun³

1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;
2. School of Computing and Information Systems, University of Tasmania, Tasmania, Australia;
3. College of Ocean Science, Shanghai Ocean University, Shanghai 201306, China

Abstract — This article introduces the design and implementation of a fishery forecasting system based on Radial Basis Function (RBF) neural network. The system was developed using the Client/Server architecture, the C# programming language in the environment of Visual Studio 2008 on the Windows7 platform. It draws knowledge from RBF neural network theory, the production historical data of pelagic fishery and the marine environment data. The system uses the Object-Oriented analysis and design method. It can quickly obtain the forecast results available to users through inputting marine environment data information and the RBF neural network model. The forecasting system includes three major functional modules, namely preprocessing fishery production data, matching production data and environmental data, training RBF neural network and making predictions. Experiments have shown that this forecasting system can generate accurate and effective pelagic fishery knowledge.

Keywords - Radial Basis Function; Fishery Forecasting; System Design

I. INTRODUCTION

After many years of development pelagic fishery has become an important part of rural fisheries development planning in China. Fishery forecast can further improve the production of pelagic fishery. In China, determining the fishery grounds quality is by means of comprehensive fishery features and marine characteristics expert knowledge. Forecasting fishery knowledge quantitation in the fishery ground is performed through using Linear Regression. For example, Professor Shen Jinao forecasted the hairtail knowledge at Chenshan in winter with establishing the prediction equations using multivariate linear regression. Professor Chen Xinjun analyzed and researched the relationship between the squid catch per unit effort and sea-surface temperature and established the linear equations. When there are two or more predicting indicators they must solve the multiple linear equations. In the same way as the development of information technology and space technology, satellite remote sensing data also plays an important role in marine fishery. For example, people can deduce the marine environmental factors related to the high fish survival using the surface scan thermal infrared data from air satellite by intelligent information processing. The marine environmental factors provide much valuable information in sea change and formation mechanism of fishery ground. At present, there are some shortcomings in forecasting the fishery knowledge. First, using multiple regression analysis method implies that the dependent

variables are independent and normally distributed. This is not easy to reach for dynamic marine environment variables. Second analyzing the marine environment factors with the method of intelligent information processing can achieve higher prediction accuracy, but these are more observational and experimental study. There are fewer business applications in fishery knowledge and fishery forecasting. The paper proposes to use the RBF neural network intelligent information processing, to provide marine environment factor input vectors closely related to the survival of fish as a means of developing a fishery forecasting system. This approach can bring the analysis of fishery grounds and fishery forecasting to business applications.

II. DESIGN'S METHODS AND THOUGH

A. Data Preprocessing

Data preprocessing can turn fishery production data and marine environmental data to the standard data for RBF network training and predicting. The excel format production data which is provided by the fisheries department, as shown in table 1, and marine environmental data needs to downloaded from the OceanWatch website, with txt format, as shown in table 2.

Table 1: The example of fishery production data

column index	column name
1	year
2	month
3	longitude
4	latitude
5	working times of vessel
6	total catch
7	average catch

Table 2: The environment data of some point

name	value
longitude	148E
latitude	42N
value of point	20.8

Each row of production data has “year”, “month”, “longitude”, “latitude”, “working times of vessel”, “total catch” and “average catch” seven columns. The marine environmental data is like a matrix. The X scale is the value of longitude, the Y scale is the value of latitude, and this

point of value is remote sense data. Each row of RBF neural network training and predicting data has “month”, “longitude”, “latitude”, “sea surface temperate”, “sea surface height”, “temperature gradient”, “habitat index” seven columns. So the standard data is a combination of production data and environmental data. In addition, the value of sea surface temperature gradient (GSST) is calculated from the value of sea surface temperature (SST) and the value of habitat index is calculated from the value of production data. Last, we need to normalize data. Given all that, this system must have a data preprocessing module.

Table 3: The example of RBF neural network training and predicting data

column index	column name
1	month
2	longitude
3	latitude
4	sea surface temperate
5	sea surface height
6	temperature gradient
7	habitat index

B. Basic Concept And Principle

Radial Basis Function (RBF) is a three layered feed-forward network as shown in Figure 1. It has a simple architecture with three layers. The input layer simply feeds the inputs to a hidden layer without any processing. The second layer is the hidden layer, which contains the radial basis functions also called the transfer function or activation functions. It determines Euclidean distance between prototype vector and input vector and then passes the result to basis function. The third layer is the output layer, which performs ordinary weighted sum of activation functions. The processing function performed in the hidden layer of the RBF network is a distinctive characteristic of the RBF neural network. The input space forms clusters of patterns. After determining the centres of these clusters, the distance from the cluster centre can be measured. The distance measure is made non-linear, so that input data closer to a cluster centre are more sensitive. Beyond this area, the sensitivity of Gaussian radial basis function sensitivity is less [2]. The concept is that this area is radially symmetrical around the cluster centre, so that the non-linear function becomes known as the radial-basis function.

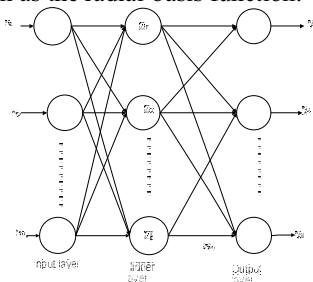


Figure 1: RBF neural network

For training Radial basis function, various approaches have been developed. Most of them can be divided into two

phases. Learning the centers and widths in the hidden layer and learning the connection weights from the hidden layer to the output layer.

For interpolation problem, The network represents the map from p-dimensional input space to the single dimensional output space which can be expressed as : Given the set of N different input points $\{x_i \in \mathbb{R}^2 \mid i=1,2,\dots,N\}$ and a corresponding target set of N real numbers $\{d_i \in \mathbb{R}^1 \mid i=1,2,\dots,N\}$, the function which satisfies the interpolation condition, $F: \mathbb{R}^N \rightarrow \mathbb{R}^1$ is

$$F(x_i = d_i), \quad i=1, 2, \dots, N \quad (1)$$

The interpolation surface has to pass through all the training data points. The approach radial basis function takes is to form a set of basis function, one for each data set. The radial basis function is of form $\phi(\|x - x_i\|)$, where $\phi(\cdot)$ is a nonlinear functions, x is the input pattern such that $x_i \in \mathbb{R}^p$, $i=1,2,\dots,N$ are the centers of the function and $\|\cdot\|$ represents a norm that is Euclidean distance. The basis function for the j^{th} hidden node is defined as follows:

$$\phi_j = \exp(-\|x - c_j\| / \sigma_j^2) \quad (2)$$

Where, σ_j is the width of the j^{th} neuron, $\|x - C_j\|$ is usually selected by the Euclidean norm of distance between the selected centers. C_j is the center of the j^{th} RBF unit.

The network output Y is obtained by a linearly weighted sum of the number of basis functions in the hidden layer. The values for the k^{th} output nodes with n number of hidden neuron can be calculated as:

$$y_k = \sum_{j=1}^n w_{jk} \phi_j \quad (3)$$

If the input datasets are as $X = \{(x_1, t_1), (x_2, t_2), \dots, (x_N, t_N)\}$, t_N is the targeted output, $y_N(X)$ is the output from hidden neuron and E is the vector of prediction errors. The sum squared output error measure can be defined as:

$$E = \frac{1}{2} \sum_{i=1}^N (y_N(x) - t_N)^2 \quad (4)$$

At the minimum of E the gradients with respect to all the weights w_{jk} will be zero and from equation (3), so

$$E = \frac{1}{2} \sum_{i=1}^N (w_{jk} \phi_j - t_N) = 0 \quad (5)$$

The above equation can be generalized as,

$$T = \phi W \quad (6)$$

The solution here is over determined in the sense that there are more rows than columns i.e.; more data points than free parameters. Therefore the weight can be calculated as:

$$T = \phi \uparrow W \quad (7)$$

Where, Φ^\dagger is the pseudo-inverse of the matrix Φ , that is;

$$\phi \uparrow = (\phi^T \phi)^{-1} \phi^T$$

Therefore, RBF neural network includes a training model and a predicting model. The predicting model is built on the basis of a model file, and the model file is formed from training after setting the number of neuron through the learning samples, when the accuracy meet the requirements. The RBF model file retains the weight and width of each neuron in the hidden layer. Usually, when meeting accuracy

requirements, in order to make faster predictions, we need to confirm the best number of neuron through constantly experiments. In addition, the trained model file in one fishery ground is not necessarily suitable for other fishery ground. So the model data is not uniquely and different fishery grounds may have different model files. After determining a good model file, we could make predictions. Given all that, this system requires to design the RBF neural network training module and predicting module.

III. SYTEM DESIGN

A. System development environment

This system is a desktop system and users operate it through installation executable files. The system development platform is Windows 7. The development environment is Visual Studio 2008, using C# language.

B. Function structure of system

The structure of the system function is designed in accordance with the methods and thoughts of study as shown in Figure 2.

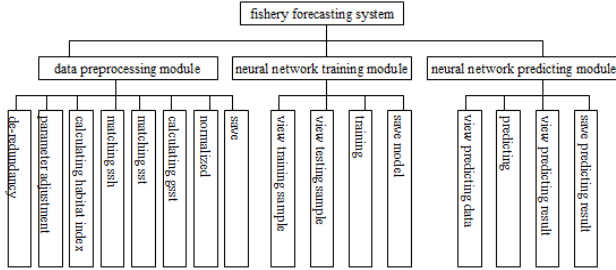


Figure 2: Function structure of system

The system is composed of three modules. They are data preprocessing module, neural network training module and neural network predicting module. These modules work by calling the necessary files, and they do not affect each other. The main functions of the modules are described below.

1) Data preprocessing module

The data preprocessing module is to get the training and predicting standard data using production data and marine environmental data after the steps of de-redundancy, parameter adjustment, calculating habitat index, matching sea surface temperature and sea surface height (SSH), calculating temperature gradient and normalized. De-redundancy is deleting row data where the value of “working times of vessel” is zero or the value of “total catch” is zero in production data. Parameter adjustment is adding the four columns “sea surface temperature”, “sea surface height”, “temperature gradient” and “habitat index” in production data. Fish habitat suitability index indicate the degree of adaptability in this sea area with the formula as follows:

$$SI_{i,j} = \frac{NET_{i,j} - NET_{i,\min}}{NET_{i,\max} - NET_{i,\min}} + \frac{C_{i,j} - C_{i,\min}}{C_{i,\max} - C_{i,\min}} \quad (8)$$

In the formula (8), SI_{ij} is habitat index in month i row j , NET_{ij} is working times of vessel in month i row j , $NET_{i,\min}$ is the smallest working time of vessel in month i , $NET_{i,\max}$ is the biggest working time of vessel in month i , C_i is average catch in month i , $C_{i,\min}$ is the smallest average catch in month i , $C_{i,\max}$ is the biggest average catch in month i . Matching sea surface temperature or sea surface height is finding the value of sea surface temperature or sea surface height and filling the production data using these values whose have the same longitude and latitude. Temperature gradient indicates the degree of temperature change in the area, formula as follows:

$$SST_{i+1,j-1} \quad SST_{i+1,j} \quad SST_{i+1,j+1}$$

$$SST_{i,j-1} \quad SST_{i,j} \quad SST_{i,j+1}$$

$$SST_{i-1,j-1} \quad SST_{i-1,j} \quad SST_{i-1,j+1}$$

Figure 3 the example of calculating GSST

$$GSST_{i,j} = \sqrt{\frac{(SST_{i,j-1} - SST_{i,j+1})^2 + (SST_{i+1,j} - SST_{i-1,j})^2}{2}} \quad (9)$$

In the formula (9), SST_{ij} is value of sea surface temperature in the point (i,j) , $GSST_{i,j}$ is the value of temperature gradient in the point (i,j) . Normalized, all data is divided into 0 to 1, formula as follows:

$$v_{i,j} = \frac{v_{i,j}}{\max_j} \quad (10)$$

In the formula (10), $v_{i,j}$ is value of row i column j , \max_j is the biggest value of column j .

2) Neural network training module

The effect of RBF neural network module is to produce model files. This file includes the number of hidden layer neuron, the width and weight of each neuron. When training, you need to import the learning sample, test sample and set the number of hidden layer neuron.

3) Neural network predicting module

The effect of RBF neural network predicting module is to produce forecasting index, the index can determine the fish abundance in fishery ground. When predicting, you need to import the standard file and model file.

IV. SYSTEM TESTING AND RESULTS ANALYSIS

A. Preparing data

Fishery production data is taken from the 2001 north pacific squid produce statistics from Shanghai Ocean University’s Fishing Technology Group. The environmental data is drawn from the OceanWatch website.

B. Testing system operation

1) Testing data preprocessing module

At the beginning interface, click on “Action” menu, select “Data preprocessing module”, open production data and marine environmental data, click each button on the interface of this module, the result shows in Figure 4:



Figure 4: Testing data preprocessing module

2) Testing neural network training module

At the beginning interface, click on “Action” menu, select “RBF neural network training module”, open learning sample and testing sample, click each button on the interface of this module, the result is shown in Figure 5.



Figure 5: Testing neural network training module

3) Testing neural network predicting module

At the beginning interface, click on “Action” menu, select “RBF neural network predicting module”, open standard file and model file, click each button on the interface of this module, the result shows in Figure 6.



Figure 6: Testing neural network predicting module

C. Relative error analysis of predicting

Setting the different number of neurons in the hidden layer to predict and calculate the relative error, the result shows in Table 4.

Table 4: The variance table

number of sample	number of neurons in hidden layer	relative error
172	20	15.24%
172	30	8.43%
172	40	7.51%
172	50	7.39%
172	60	7.21%

We can see from Table 4 that with the increase in the number of hidden layer neurons, the relative error gradually decreases. The relative error trends to stabilize when the number of neurons increases to 40 onwards. Therefore, considering the efficiency of prediction and forecasting accuracy, such samples should have 40 neurons selected.

V. APPLICATION AND PROSPECT

After several rounds of improvement, the standardization and versatility of the system are further enhanced. Currently, the system has been applied to the fishing technology group of Shanghai Ocean University.

Actual operation proves that the system is reliable and produces good performance. Forecasting by this system using marine environmental factors and fishery production factors not only improves the forecasting accuracy, but also results in business applications. Forecasting reports are published in the GIS Remote Sensing Laboratory and Marine Fisheries Website of Shanghai Ocean University. Users can download and use the reports in production practices. Consequently this system has good application prospects.

ACKNOWLEDGMENT

This work is supported by Shanghai Education Committee (Grant No.09YZ267); State 863 projects (Grant No.2007AA092201) and National Development and Reform Commission special (Grant No.2060403).

REFERENCES

- [1] Yang Yebo, Yuan Hongchun, A forecasting model of WCPO Katsuwonus Pelamis Purse Seine Catch Based on Rough-SVM. Chengdu, China: International Conference on Intelligent Systems and knowledge Engineering. 2007.10.(in Chinese)
- [2] Abraham, Ajith. Artificial Neural Networks. Handbook of Measuring System Design. : John Wiley & Sons, Ltd, 2005.
- [3] Fan Wei, A study on application of satellite remote sensing in marine fishing-ground analysis and fishing condition forecasting, Doctor Degree thesis. Shanghai, China: east China normal university. 2005.7.(in Chinese)
- [4] Su Fenzhen, Zhou Chenghu, Shi Wenzhong, Rough-Set-Based spatiotemporal assignment mining model with its application for marine fishery. Journal of remote sensing. 2005.4(9):339-403.
- [5] Tian Siquan, Chen Xinjun, Impacts of different calculating methods for nominal CPUE on CPUE standardization. Journal of Shanghai ocean university. 2010.3(19):240-244.
- [6] Chen Xinjun, Liu Bilin, Tian Siquan, Forecasting fishing ground of Ommastrephes Beltrami by using a sst-based habitat suitability model in the northwest pacific. Transactions of Oceanology and Limnology.2010(3):41-45.