ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 15 (2011) 3240 – 3244

Procedia Engineering

www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Prediction of Fishing Ground based on RBF Neural Network

Hongchun Yuan[a], Xiaoqian Shen[a], Xinjun Chen[b, a]*

[a]*College of Information Technology, Shanghai Ocean University, 999 Hucheng Huan Road,201306,Shanghai, China*
[b]*College of Ocean Science, Shanghai Ocean University,999 Hucheng Huan Road, 201306, Shanghai, China*

Abstract

This study tested Radial Basis Function Neural Network (RBFN) as an intelligent method to fulfill the prediction of fishery forecasting in Southwest Atlantic on Illex argentines. Due to the existing drawback of fuzzy C-means (FCM) RBF which is time consuming, we used symmetry-based Fuzzy C-means (SFCM) to improve the effectiveness of RBF. Altogether Six marine environmental factors are considered which are months, longitude and latitude, sea surface temperature (SST), Sea surface Height (SSH) and chlorophyll for predicting the Habitat Suitability Index (HSI). The traditional calculation methods of HSI are statistical ways such as multiple linear regressions. The results obtained from the SFCM/RBF model were compared with Multiple Linear Regressions in terms of accuracy criterions MSE, RMSE. Through the prototype system, it is shown that the intelligent model has high predictive ability and better goodness of fit compared with statistical models.

Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords:Radial Basis Function Neural Network; Fishery forecasting; Fuzzy C-Means; Point Symmetry Distance; Multi-linear regression; Habitat Suitability Index;

1. Introduction

Forecasting plays a significant role in fishery management. Environmental factors are used, in fishery operation, to improve the forecasting results. Since the early 80s of last century, habitat suitability index (HSI) models have been proposed to study the distribution of fish habitat and fisheries forecasting. The majority calculating methods are in statistical way, which includes linear regression and multi-linear

* Corresponding author, Tel.: +021-61900623.
E-mail address: hcyuan@shou.edu.cn,.

doi:10.1016/j.proeng.2011.08.608

regression. Wang [1] used normalized CPUE as a habitat suitability index and the environmental factor as single factor to predict the Indian Ocean tuna habitats. Feng [2] used quantile regression to establish the Argentine Illex habitat model. But there are many problems in HSI model. The existing habitat model does not consider the interaction between the factors. But, in practice, mutual influence occurred by a variety of factors together is inevitable. RBFN is outstanding because of its simple structure and fast learning, but forecasting the production of the Argentine Illex is still blank. Due to amounts of marine environmental data, RBF network will take a period of time and may produce a high error. Even though a combination of clustering methods in RBF networks has been proven by Sarimveis et al. [3] to be fast in training, it still produces a larger error. For the standard clustering algorithms, it still lacks the ability to choose the most accurate and informative centers.

In this study, the applicability of RBFN for modeling the HSI of the Southwest Atlantic on Illex argentines is investigated. An improved network combining RBF with SFCM will be implemented in HSI model and the other processes of HSI creation will be maintained, ending with performances of this networks were compared with a traditional HSI model.

2. Theory and Methods

2.1. Fuzzy C-Means(FCM) Clustering Algorithm and the Point Symmetry Distance

FCM-based algorithms are the most widely used fuzzy clustering algorithms in practice. The basic FCM algorithm is to divide N vectors x_i (i = 1,2, ..., n) into c fuzzy groups and find each of the cluster center, aiming to minimize the objective function. In the basic FCM, the similarity measure between the patterns or objects is defined by the Euclidean distance, which could cause hyper spherical-shaped clusters, failing to detect clusters.

The point symmetry distance was used for training single-layer neural networks to cluster data is, a new competitive learning algorithm, proposed in [4]. Given N patterns, x_i (j=1,…, N) and a cluster center c, the point symmetry distance between a pattern and the reference vector c is defined as

$$Distance_s = \min_{\substack{i=1,\dots,N \\ and\, i \neq j}} \frac{\left\| (\ddot{x}_j - \ddot{c}) + (\ddot{x}_i - \ddot{c}) \right\|}{\left(\left\| (\ddot{x}_j - \ddot{c}) \right\| + \left\| (\ddot{x}_i - \ddot{c}) \right\| \right)} \tag{1}$$

2.2. An improved neural network model combining SFCM and RBF

A radial basis function network has a three layered feed-forward structure consisting of a single hidden layer for a given number of locally tuned units which are fully interconnected to an output layer of linear units. The processing function performed in hidden layer of the RBF network is distinctive characteristic of RBF neural network. For the hidden layer, this study uses the proposed algorithm for selecting the most significant input centers, based on the symmetry based Fuzzy C-Means clustering method. The SFCM [5] approach adopted in the present study can be simply listed in the algorithm below:

- Initialization. Randomly choose K data points from the data set to initialize K cluster centers,

$$\ddot{c}_r = (\ddot{c}_{r1}, \dots, \ddot{c}_{rm})^T \quad , r=1,2\dots,K \tag{2}$$

- Fine-Tuning. For pattern x_i , find the cluster center which is nearest in the symmetrical sense. That is, find the cluster center k that is nearest to the input pattern using the minimum-value criterion:

$$k = Arg \min_{r=1\dots K} d_s(\ddot{x}_j, \ddot{c}_r) \tag{3}$$

where the point symmetry distance is computed by (1). If the point symmetry distance is smaller than a pre-specified parameter θ , then we update the membership u_{ij}

$$\begin{cases} u_{ij} = 1, & if\ i = k \\ u_{ij} = 0, & if\ i \neq k \end{cases}$$

$$u_{ij} = \frac{1}{\sum_{r \in S_r(t)}^{c} \left(\frac{d_{ij}}{d_{rj}}\right)^2} \tag{4}$$

- Updating. Compute the new centers of the k clusters using (3). The updating rule is given below, where $S_k(t)$ is the set whose elements are the patterns assigned to the kth cluster at time t.

$$c_r(t+1) = \frac{\sum_{i \in S_r(t)} u_{ij} \dddot{x}_j}{\sum_{i \in S_r(t)} u_{ij}} \tag{5}$$

- Continuation. If there are no patterns change categories, or the number of iterations has reached a pre-specified maximum number, then stop. Otherwise, go to Step 2.

The basic idea of the SFCM algorithm is very simple to understand. SFCM algorithm is the modified version of FCM algorithm, but the modification is only done in Step 2 and Step 4 of FCM algorithm while the rest of the steps of SFCM algorithm remain the same as FCM algorithm. In this study, the weights of RBF network based on Least Square Method (LMS).

2.3. Traditional HSI Modeling Approaches

Habitat suitability index (HSI) model is firstly presented by the U.S. Fish and Wildlife Conservation Commission in order to quantitatively describe the quality of wildlife habitat. Model assumes a positive correlation between the index and habitat environment, while the index higher, the better quality of habitat. Generally it uses 0.0 to 1.0 to scale the impact of a habitat suitability index model. A Multi-linear Regression (MLR) model is defined by the equation:

$$Y_i = \beta_0 + \sum_{i=1}^{N} \beta_i X_i \qquad \text{i=1,2,…,n} \tag{6}$$

Table 1.Coefficient for multiple regression models

Intercept	Coefficients for month	longitude	latitude	SSH	SST	Chl.
Y	a_1	a_2	a_3	a_4	a_5	a_6
5.492	0.245	2.226	2.242	0.420	0.560	0.303

Where: Y is the response (Habitat Suitability Index) in the ith case, ai is the value of the independent variable in the ith case, β_0 and β_i are the regression coefficients. Table 1 represents the coefficient of multiple regression models.

3. Experiments and results

3.1. Implementation of Prototype System

The data conducted in this study is divided to 2 parts: production statistics data and marine environmental data. Data from 2000 to 2003 will be the train data (70%) and test data (30%), while data in 2004 will be the benchmark for prediction. There are two parts in the system. First part is neural network training module, where we could train the neural network by SFCM/RBF and standard RBF methods. At second part, predicting module will validate the effectiveness of three different methods.

Fig.1: (a) neural network training module; (b) neural network predicting module

3.2. Results and Comparative Study

The essential part of RBFN algorithm is to choose the RBF hidden neurons. The set of raw data by year is about 50 to 130. In order to predict the fishing ground of some unknown year, we will use the data covering several years, thus the amount of the training data would be increased according to the needs. In this study, the amount of training data is around 400 to 700. Hence, we use 459 sets of data respectively test the suitable number of hidden neurons on SFCM/RBF network and standard RBF. The adopted error criterion is the normalized root mean squared error (RMSE).

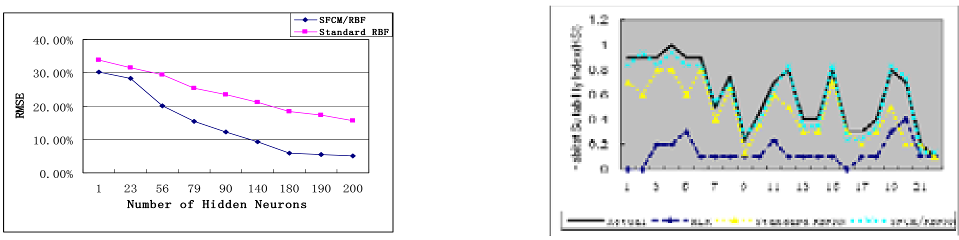

Fig.2: (a) Comparison of training performance between SFCM/RBF, RBF; (b) Comparison of prediction result between SFCM/RBF, RBF, MLR

It is observed from the chart that the RMSE is declining with the increasing of the hidden neurons, and the RMSE of SFCM/RBF method represents more obviously dropping trend than RBF method. The curve of SFCM/RBF is not stable until 180 hidden units, but the accuracy is higher than RBF method, of which

RMSE is still falling. Hence, with the same number of hidden neurons, the SFCM/RBF method has better performance. Compared with RBF and MLR methods, SFCM/RBF could better reflect the actual situation.

Table 2: MSE values (in %) for predicting Habitat Index value

	SFCM/RBF	RBF	Multi- Linear Regression
MSE	0.027	0.037	0.069
CPU	0.3 s	0.4 s	45.1 s

Many different criteria are used to judge the network performances. Generally, the smaller the MAE values, the smaller error between predicted and actual data. The network training speed of the SFCM RBF model is much faster than that of the standard RBF model and MLR, i.e., the average Central Processing Unit (CPU) times needed for training take 0.3 s for the SFCM /RBF model and 0.4 s for the standard RBF while 45.1 s for MLR. Therefore, it can be concluded that, under the same conditions, the SFCM/RBF network exhibits better properties than standard RBF and MLR.

4. Conclusions

In this study, the intelligence method was presented and used to optimize the output variables of Habitat Suitability Index implemented for fishing ground prediction. The relative environmental factors of Southwest Atlantic Illex argentinus were used as a test to confirm the correction and accuracy of the proposed network. Through the comparison with simple RBF and MLR method, it shows that the SFCM/RBF neural network structure can provide prediction of fishery ground with acceptable accuracy. It is observed that SFCM/RBF network need as much less considerable amount of time. Alternatively, the proposed approach can be used to predict other kind of fish HSI with a little adaptation in the number of input variables. More studies are needed to apply the developed system to other fish forecasting areas and to assess the degradation of fish forecasting predictions due to inherent errors in the measured database.

Acknowledgment

This work is supported by Shanghai Education Committee (Grant No. 12ZZ162); State 863 projects (Grant No.2007AA092202, No.2007AA092201) and National Development and Reform Commission special(Grant No. 2060403).

References

[1] Wang Qiao. Bigeye tuna habitat suitability index model. *Shanghai Fisheries University*, Shanghai;2006. p. 1-40.

[2] Feng bo,The habitat suitability index of illex argentinus by using Quantile regression method in the Southwest Atlantic.*Ocean Bulletin Press* 2010;**1**:15-22.

[3] H. Sarimveis. A fast training algorithm for RBF networks based on subtractive clustering. *Neurocomputing* 2003; **51**:501-505.

[4] M.C.Su. A Competitive Learning Algorithm Using Symmetry. *Communications and Computer Sciences*; 1999, p. 680-687.

[5] Lim Eng Aik, Modified Clustering Algorithms for Radial Basis Function Networks. *Master Dissertation*,Malaysia:Pulau Pinang; 2006.