A Forecasting Model of WCPO *Katsuwonus Pelamis* Purse Seine Catch Based on Rough-SVR#

Yebo Yang [1] Hongchun Yuan [1] Ying Chen [2]

[1]College of Information Technology, Shanghai Fisheries University, China
[2]School of Information Systems, University of Tasmania, Australia

Abstract

Fishery prediction by using long-term accumulated hydrology elements and catch statistic data has been an urgent requirement in aquatic domain. In this paper, 18 different WCPO (Western and Central Pacific Ocean) hydrology elements were collected and a key influence element set was found by using Rough Set Theory before training to build up the forecasting model rather than using ε-SVR directly. Comparative experiments with traditional ε-SVR show that the Rough Set could remove redundancy elements effectively; the Rough-SVR results in a better goodness of fit than the traditional ε-SVR and it is superior to multiple regression analysis.

Keywords: Rough Set; Support Vector Regression; Katsuwonus Pelamis; Predicting of the purse seine catch

1. Introduction

Tuna fishery is one of the most important developing fields during the Eleventh Five-Year Plan which has attracted many Chinese experts in fisheries into the study of fishery prediction. Traditional predicting methods adopt a multiple regression analysis [1] while dynamic hydrology elements are not prone to meet its requirement of input, and don't produce a high goodness of fit in results. FN (Functional Networks) and ANN (Artificial Neural Networks) were also adopted [2], while the problem of local optima can't be solved thoroughly. In the face of long-term accumulated hydrology elements and catch statistic data of WCPO (Western and Central Pacific Ocean), there emerged an urgent requirement to make the fishery prediction more precise by using hydrology elements such as the Sea Surface Temperature (SST), Sea Surface Temperature Anomaly (SSTA) and sea temperature of different depth and also adopting a more appropriate data mining method.

Support Vector Machine (SVM) is a new machine learning method introduced by Vapnik [3] which implements the structural risk minimization inductive principle with the purpose of obtaining a good generalization from limited size data sets. It has numerous attractive features and promising empirical performance. It provides high generalization performance without the need to add a priori knowledge, even when the dimension of the input space is very high. This ability results from their main difference from the other types of neural networks, that they are an exact implementation of the Structural Risk Minimization (SRM) principle.

Attribute simplifying is one of the most successful applications of the Rough Set Theory [4]-[5]. For many of the large-scale systems, only a part of the data table attributes need to be reserved. It will improve the clarity of potential knowledge if the redundant attributes could get removed. It turned out to have a better performance that the attributes of training sets have been simplified before training.

In Section 2, we introduce the concept and algorithm of Rough Sets – Support Vector Machine as the discernibility matrix proposed by Bazan, Skowron & Synak [6]. It is adopted to remove the redundant attributes. Section 3 introduces the modeling processing of WCPO *katsuwonus pelamis* purse seine catch by using hydrology temperature elements （20°N～25°S，175°W～0°W）, purse seine net number and amount of catch. Finally the paper will present the evaluation and result of applying this method for prediction as compared with the traditional Support Vector Regression (SVR).

2. Rough Sets and Discernibility matrix

2.1. Rough sets

An information system is a pair S = (U,A), where U is the universe of discourse with a finite number of objects (or entities). A is a set of attributes defined on U. Each a∈A corresponds to the function a : U → Va, where Va is called the value set of a. Elements of U are called situation, objects or rows, interpreted as, e.g., cases, states, patients, observations.

With any subset of attributes B⊆A, we associate the information set for any object x ∈ U by

InfB(x) = {(a, a(x)): a ∈ B}

An equivalence relation called B-indiscernible relation is defined by

IND (B) = {(x, y) ∈ U × U: InfB(x) = InfB(y)}

Two objects x, y satisfying the relation IND(B) are indiscernible by attributes from B. [x]B is referred to as the equivalence class of IND(B) defined by x. A minimal subset B of A such that IND(B) = IND(A) is called a reduct of S. Suppose S = (U,A) is an information system, B ⊆ A is a subset of attributes, and X ⊆ U is a subset of discourse, the sets

B(X) = {x ∈ U: [x]B ⊆ X},B(X) = {x ∈ U : [x]B_X _= φ}

are called B-lower approximation and B-upper approximation respectively.

In a decision table DT = (U,A ∪ {d}), where {d} ∩ A = φ, for each x ∈ U, if [x]A ⊆ [x]{d}, then the decision table is consistent, or else it is inconsistent.

2.2. Discernibility matrix

Given a decision table DT = (U,A∪{d}), where U = {u_1,u_2,…,u_n}, A = {a_1,a_2, …,a_k}, by discernibility matrix of the decision table DT we mean the(n × n) matrix:

$$M(DT) = [C_{i,j}]_{i,j=1}^{n}$$

such that Ci,j is the set of attributes discerning u_i and u_j . Formally:

$$C_{i,j} = \begin{cases} \{a_m \in A : a_m(u_i) \neq a_m(u_j)\} & If\ d(u_i) \neq d(u_j) \\ \phi & otherwise \end{cases}$$

The discernibility function corresponding to M(DT) is defined as follows:

$$f(DT) = \wedge_{i,j}(\vee C_{i,j}), C_{i,j} \neq \phi$$

3. Support Vector Regression

The model for regression as following: given a training set S={(x_i， y_i)} (i =1,2,3,…,m，(x_i， y_i)∈ R^n·R) of input x_i and associated targets y_i , the goal of regression problem is to fit a flat function *f(x)* which approximates the relation inherited between the data set points and it can be used later on to infer the output y for a new input data point x .

To predict the catches of *Katsuwonus Pelamis*, x_i is the elements that affects the catch e.g. hydrology elements and purse seine net number, n is the number of the elements, $y_i \in R$ is the number of input data, here is the catch data, m is the number of input data in this article,.

Suppose the function *f(x)* is expressed as:

$$f(x) = <\omega, \phi(x)> + b \quad \phi : R^n \to F, \omega \in F, b \in R \quad (1)$$

where <,> is dot product of vector, b is a bias term, and ϕ is a nonlinear map which mapping the input x into a high-dimensional feature space F

According to SRM principle, that function *f(x)* is flat in the case of Eq. (1) means that one seeks the minimization of the following expression:

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} L(f(x_i), y_i) \quad (2)$$

where L(・) is a loss function, C is a constant.

Many forms for the loss function can be found in existing literature: e.g. linear, huber and quadratic loss function, etc. In this paper, Vapnik's loss function [5] is used, which is known as ε -insensitive loss function and defined as:

$$L(y, f(x)) = \begin{cases} 0 & |f(x) - y| < \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases} \quad (3)$$

Thus, the regression problem can be written as a convex optimization problem:

$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \quad (4)$$

$$\text{s.t} \begin{cases} y_i - <\omega, \phi(x_i)> - b \leq \varepsilon + \xi_i \\ <\omega, \phi(x_i)> + b - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

ε>0 is a predefined constant which controls the noise tolerance, the constant C>0 determines the trade-off between the flatness of f and the amount of tolerable deviations, which is larger than ε.

Through introducing a Lagrange function, the optimization problem (4) and (5) can be solved in their dual formulation, which is expressed as follows:

maximize

$$-\frac{1}{2}\sum_{i,j=1}^{m}(\alpha_i + \alpha_i^*)(\alpha_j - \alpha_j^*) < \phi(x_i), \phi(x_j) > \quad (6)$$
$$-\varepsilon\sum_{i=1}^{m}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{m} y_i(\alpha_i - \alpha_i^*)$$

$$\text{s.t} \begin{cases} \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (7)$$

The optimal value of α_i , α_i^* can be obtained by solving the dual problem (6), (7), accordingly, the ω can be described by:

$$\omega = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i) \quad (8)$$

thus, $f(x)=\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)<\phi(x_i),\phi(x)>+b$ (9)

and the value of b can be computed according to the Karush-Kuhn-Tucker (KKT) conditions. Equation (9) is so-called support vector machines regression expansion.

It can be seen clearly from Eq. (9) that we only need the dot product of input data instead of computing the value of ω and $\phi(x)$. We introduce kernel instead of nonlinear mapping, i.e. $K(x,x')=<\phi(x),\phi(x')>$ then Eq. (9) is rewritten as follows:

$$f(x)=\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)K(x_i,x)+b \qquad (10)$$

where kernel K(x, x') are arbitrary symmetric functions, which satisfy the Mercer condition.

4. Modeling Processing

Since the dimension of hydrology elements could be 18 yet only concerns the attributes of ocean temperature, more than 70 concerns the ocean temperature, salinity and two velocities of flow (vertical and horizontal), the Rough Set Theory was applied to remove the redundant attributes.

The problem of the purse seine catch prediction may be thought as a multistage process as shown in Fig. 1.

Preprocessing is to process the data before training, including data rectifying, filling up missing data, data alignment and so on.

Attribute reduction removes the redundant attributes by rough set.

Organizing training data set should be corresponding to different situations in the domain of fisheries, basically four types of situations: La Nina year, El Nino year, Strong El Nino year and Normal year.

Regression is the modeling of the WCPO *katsuwonus pelamis* purse seine catch by ε-SVR to get the prediction model, meanwhile testifying the model and feedback to the ε-SVR method to adjust parameters and kernel function to get the optimized result.

4.1. Data sources and data formats

Because the development of WCPO Tuna fisheries is steady-going during the years from 1984 to 2003, there are no dramatic movements in the scope of catch and the amount of catch. This paper adopts the data from 1995 to 2000 as training data and the data of the first half of 2001 as prediction data.

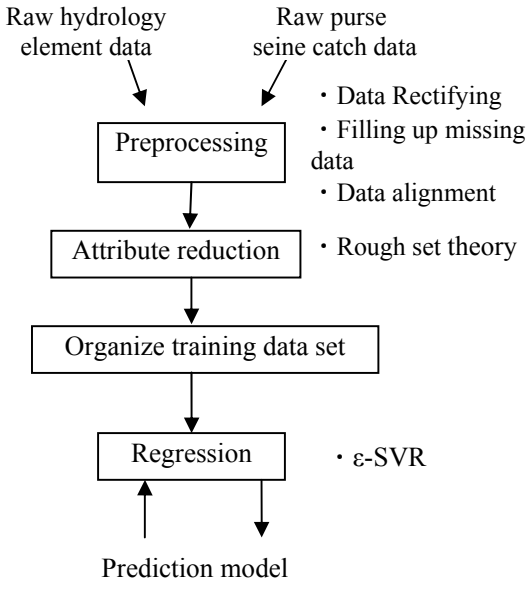

Fig. 1: The stages comprising the regression problem.

The Catch & Effort (C&E) data of WCPO purse seine tuna catch was provided by South Pacific Ocean Forum Fisheries Committee (FFC) from January 1990 to July 2001. A part of the WCPO hydrology temperature element data was downloaded from IRI/LDEO Climate Data Library (http://iridl.ldeo.columbia.edu, data from 1980 to now) and was recomposed to restore into an SQL database. Another part of data was computed as the difference in temperature per meter between two depths, based on the original hydrology data under the suggestion of domain experts. Table 1, Table 2 and Table 3 below show the data formats in detail.

4.2. Preprocessing

As the spatial grid area of *Katsuwonus Pelamis* Purse Seine Catch is 5°×5°, the grid area of training data should be 5°×5° also. But the spatial grid area types of hydrology temperature elements are 1°×1° and 1.875°×1.875°, the data must be transformed to fit into the 5°×5° grid uniformly. We adopted arithmetic average method to preprocess there data**,** arithmetic average function:

$$T(m,n)=\frac{\sum T(i,j)}{N}$$

$$(m-2.5\le i\le m+2.5,\ n-2.5\le j\le n+2.5) \qquad (11)$$

where T(m, n) is the average attribute value(SST, SSTA or others) in the 5° area which takes point (m,

n) as center (m is longitude, n is latitude). T(i, j) is the attribute values in the 5° area. N is the sum of attribute number.

Attribute Name	Spatial grid area
Effort (Purse Seine Net Number)	Type A*
Katsuwonus Pelamis Purse Seine Catch	Type A

Table 1: Catch & Effort (C&E) Data provided by South Pacific Ocean Forum Fisheries Committee (FFC)

Difference in temperature per meter between two depths	Spatial grid area
12.5m depth and ocean surface	Type B
37.5m depth and 12.5m depth	Type B
62.5m depth and 37.5m depth	Type B*
87.5m depth and 62.5m depth	Type B
137.5m depth and 87.5m depth	Type B
187.5m depth and 137.5m depth	Type B
237.5m depth and 187.5m depth	Type B
287.5m depth and 237.5m depth	Type B

Table 2: Hydrology temperature element data computed under the suggestion of domain experts

Attribute Name	Spatial grid area
Sea Surface Temperature(SST)	Type C*
Sea surface temperature anomaly (SSTA)	Type C
Ocean Temperature of 12.5m depth	Type B*
Ocean Temperature of 37.5m depth	Type B
Ocean Temperature of 62.5m depth	Type B
Ocean Temperature of 87.5m depth	Type B
Ocean Temperature of 137.5m depth	Type B
Ocean Temperature of 187.5m depth	Type B
Ocean Temperature of 237.5m depth	Type B
Ocean Temperature of 287.5m depth	Type B

Table 3: Hydrology temperature element data downloaded from IRI/LDEO Climate Data Library

(Type A is 5°×5°, Type B* is 1.875°×1.875°, Type C* is 1°×1°)

4.3. Attribute reduction

Due to the complex operation resulting from a large amount of data, we picked the hydrology temperature element data of year 2001 to be tested to find out the key attributes and removed the redundant attributes. The result is shown in Table 4.

4.4. Modeling by ε-SVR

The training data sets were picked and organized from hydrology temperature element data and Catch & Effort database from Jan 1995 to Jun 2000 which yielded 2867 data in total. There are 13 attributes in it including 12 hydrology temperature elements and 1 effort element. The predict data sets were organized as 151 data from the first half of 2001.

Sea Surface Temperature(SST)	Ocean Temperature of 187.5m depth
Sea surface temperature anomaly (SSTA)	Ocean Temperature of 237.5m depth
Ocean Temperature of 12.5m depth	Ocean Temperature of 287.5m depth
Ocean Temperature of 62.5m depth	Difference in temperature per meter between 62.5m depth and 37.5m depth
Ocean Temperature of 87.5m depth	Difference in temperature per meter between 137.5m depth and 87.5m depth
Ocean Temperature of 137.5m depth	Difference in temperature per meter between 287.5m depth and 237.5m depth

Table 4: Key attributes of hydrology temperature elements found by discernibility matrix.

The conventional multiple regression analysis, ε-SVR method with 19 attributes and the Rough ε-SVR method with 13 attributes were applied to the data sets. In order to evaluate the accuracy of the three methods, we introduced the R^2, the goodness of fit indicator. R^2 is a number between 0 and 1，the closer it gets to 1 the better goodness of fit is indicated. The function of R^2 is:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (Y - Y')^2}{\sum (Y - \bar{Y})^2} \quad ; \qquad (12)$$

where $$SSR = \sum (Y' - \bar{Y})^2 \quad ; \qquad (13)$$

$$SSE = \sum (Y - Y')^2 \quad ; \qquad (14)$$

$$SST = SSR + SSE \text{ 。} \qquad (15)$$

SSR is regression sum of square; SSE is sum of square of residues; SST is total sum of square; Y is the actual catch; Y' is predicted catch; $\bar{Y}$ is the average value of data set Y.

The goodness of fit of the three approaches was measured, and the results are given in Table 5 indicating a better output for Rough ε-SVR. The Rough ε-SVR can predict the data more precisely with less attributes. All redundant attributes have been removed effectively and the result is desirable.

Method	Attributes involved	R2
Multiple regression analysis	19	0.51
ε-SVR	19	0.8149
Rough ε-SVR	13	0.8261

Table 5: Goodness of fit: Rough ε-SVR method vs. multiple regression analysis and conventional ε-SVR method.

The prediction curves are shown in Fig. 2, 3 and 4 below.

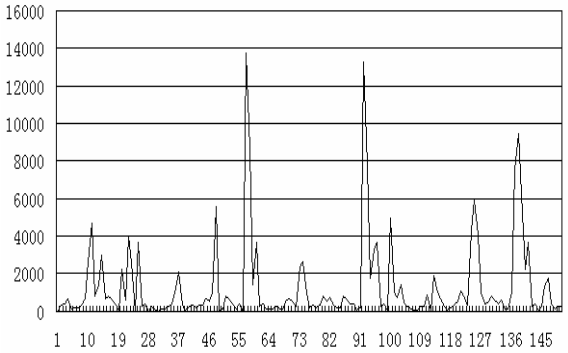

Fig. 2 Actual catch of the first half of 2001

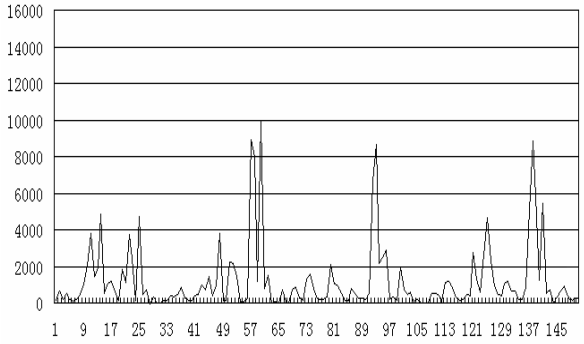

Fig. 3 Prediction curve by conventional ε-SVR method with 19 attributes

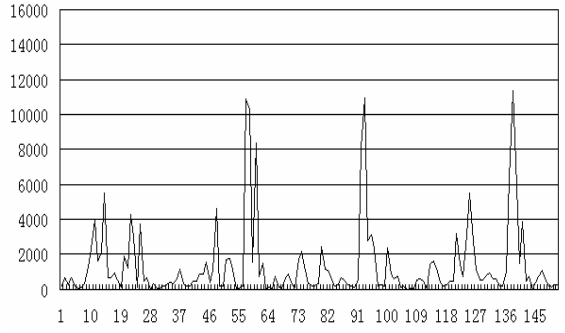

Fig. 4 Prediction curve by Rough ε-SVR method with 13 attributes

4.5. Kernel and parameter selection

There are a great number of factors that influence the regression effect. To different kernel functions the goodness of fit is shown in Table 6. The results of four classifiers with different kernel function are shown: linear kernel (LK), a polynomial kernel (PK), Gaussian radial basis function kernel (RBF) and sigmoid kernel (SK).

SVM Kernel	LK	PK	RBF	SK
Parameter	NO	d=2	σ=0.001	ν=0.001
Goodness of fit	0.7584	0.8261	0.7520	0.2562

Table 6: The influence of kernel function

The experimental results show that the goodness of fit is the highest with polynomial kernel (*d*=2), where d is the rank of polynomial. Namely

$$K(x_i,x_j) = (\gamma x_i^T x_j + r)^2 \qquad (16)$$

The choice of γ and C appeared not having too much inference on the results as they got a difference no more than 0.01.

5. Conclusions

As a result of this research activity, the integration of Rough Set Theory (RST) and Support Vector Regression (SVR) can predict the purse seine catch of *Katsuwonus Pelamis* satisfactorily. It exhibits desirable goodness of fit. It represents a more precise result with fewer attributes than the method with the redundant ones.

However the limitation of the research lies in the fact that only 18 hydrology temperature elements were used. Our next research is to involve more hydrology elements such as salinity, two velocities of flow (vertical and horizontal) and Sea Surface Height (SSH) etc.

Another key issue identified in the research is the way training data sets are organized and the model selection of Support Vector Machine. Our next research will also include the experiments on different situations such as La Nina year, El Nino year, Strong El Nino year and Normal year and select a more suitable model for them in order to reach a better prediction result after training.

Acknowledgement

This work was supported by Shanghai Municipal Education Commission (06KZ016).

References

[1] L.P. Xu, The prediction of WCPO *Katsuwonus Pelamis* Catches (in Chinese). *Master thesis of Shanghai fisheries university*, 2006.

[2] A. Iglesias, B. Arcay, J.M. Cotos, J.A. Taboada, C. Dafonte. A Comparison Between Functional Networks and Artificial Neural Networks for the Prediction of Fishing Catches. *Neural Computing & Applications,* pp.24-31, 2004.

[3] V. N.Vapnik, *Statistical Learning Theory*. John Wiley & Sons, New Work 1998.

[4] Synak, Piotr. Temporal Templates and Analysis of Time Related Data, Rough Sets and Current Trends in Computing, *Second International Conference, RSCTC 2000 Banff,* Canada, October 16-19*,* 2000. Revised Papers

[5] G. Y. Wang, Rough Set Theory and Knowledge Acquisition. *Xi'an Jiaotong University Press*, Xi'An, China 2001.

[6] J.G. Bazan, , A.Skowron, , P. Synak, Discovery of Decision Rules from Experimental Data. *Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC'97).* San Jose: San Jose State Univ, pp.526-533, 1994.