Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 15	No. 1	pp. 337-342	2016

Original Research Paper

Research on Application of Support Vector Machine Method in Ningbo Marine Ecological Environment Security Prediction

Zhang Wei-peng
School of Mechanical and Electronic Engineering, Ningbo Dahongying University, Ningbo, 315000, China

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 12-10-2015
Accepted: 18-11-2015

Key Words:
Ecological environment security
Ningbo sea
Support vector machine

ABSTRACT

Using the data related to the marine ecological environment of Ningbo City, China from 2003 to 2013, an index system for the prediction of Ningbo marine ecological security was established from three aspects, pressure, state and response. Support vector machine (SVM) method was adopted to establish the prediction model for Ningbo marine ecological environment security. Then phase space reconstruction was performed on multivariate time series, and the evolution trend of Ningbo marine ecological environment security from 2014 to 2016 was predicted. The result showed the SVM prediction model had a satisfying simulation accuracy. It was predicted that the percentage of Ningbo sea areas with water quality of worse than Grade IV from 2014 to 2016 was about 50%, indicating an unsound marine environmental security in Ningbo.

INTRODUCTION

With the rapid development of marine economy in Ningbo City, China speeds up the resource consumption and the damage on the ecological environment. The main manifestations include an obvious increase of sea pollutants, gradual expansion of offshore pollution scope, obvious aggravation of seawater eutrophication in the coastal areas, frequent occurrence of pollution accidents such as red tide, oil spilling and illegal disposal, as well as the aggravation of marine natural and ecological damage. Nearly 70% of the Ningbo sea area has been severely polluted (Chen 2012, Dai et al. 2013). The increasingly exhausted marine resources and severely polluted marine environment, not only influence and restrict the sustainable development of marine economy, but also will certainly imperil the marine space and base on which the people's survival and development depend in the coastal area. It has become the most emergent problem for comprehensive marine economic development on how to realize the win-win strategic goal of environment and economy required for a sustainable development.

According to the structural risk minimization principle proposed by Vapnick (Cortes et al. 1998), support vector machine (SVM) improves the generalization capability of learning machine greatly. The small errors obtained from a limited training sample set are capable of ensuring small errors for independent test sets. SVM has a stronger theory basis and better generalization performance than the neural network learning algorithm based on empirical risk minimization principle. In addition, as a convex quadratic optimization problem, the SVM algorithm ensures that extremal solution is found out as a globally optimal solution (Raudys et al. 2000, Burges 1998). SVM is capable of coping with actual problems such as small sample size, nonlinearity, high dimension and local minimum problem. To a large extent it solves the problems such as model selection and over-fitting, nonlinearity and dimensionality disaster, and local minimum. Therefore, SVM is becoming a new research hotspot in the machine learning field after pattern recognition and neural network (Ma et al. 2003).

This article selected the natural, social and economic indices, which can reflect the level of Ningbo marine ecological security, establish an index system for Ningbo marine ecological security prediction based on "PSR" (pressure-state-response) model. Using the SVM's advantages in handling problems of topological structure and small sample size and in generalization capability, a SVM-based prediction model for Ningbo marine ecological environment security was established. This work provides methodology and basis for maintaining and recovering the security and stability of the marine ecological system and sustainable social and economic development in the Ningbo sea area.

ESTABLISHMENT OF SVM-BASED PREDICTION MODEL FOR NINGBO MARINE ECOLOGICAL ENVIRONMENT SECURITY

Establishment of index system of Ningbo marine ecological environment security prediction: The establishment of an index system of marine ecological environment security prediction which can reflect the features of the Ningbo sea area and a scientific prediction model for marine ecological environment security is very important for understanding the marine ecological security state and its evolution trend in this area. We adopted the "PSR" model (Zhang et al. 2014, Zhang et al. 2009) proposed by European Environment Agency. Considering the deep influence of social and economic activities on the marine ecological environment system, actual situations in Ningbo City and data availability, three subsystems were selected for the index system. The specific idea was as follows: Human social and economic activities and natural disasters increase the interruption on ecological system; the resulting pressure (P) forces the state of ecological system state (S) to change; the state change will influence human beings or ecological system and urge human beings to directly or indirectly respond (R); in turn the response may exert an effect on the pressure, state or the influence due to state change, so as to make the feedback keep stable and balanced. Based on this idea, an index system of marine ecological security evaluation was established, as given in Table 1.

Data collection: Project members collected the Ningbo marine environment data in 2003-2013 from Ningbo Statistic Yearbook, which are depicted in Table 2.

Standardization of Ningbo marine ecological security prediction indices: In order to eliminate the influence of the dimension of indices, the project adopted min-max normalization method for data standardization (Ling et al. 2013). There was a positive/negative correlation between each index and ecological security index. For the index with positive correlation, equation (1) was used for standardization; for that with negative correlation, equation (2) was used.

Table 1: Index system for ecological environment security prediction in Ningbo sea area.

Type of Index	Content of Index	Value
Pressure	Natural Population Growth Rate (x/1000)	X1
	GDP Per Capita (RMB)	X2
	Total Number of People (10,000)	X3
	Sea Area Per Capita (m^2)	X4
	Port Throughput (10kt)	X5
	Total Discharge Amount of Industrial Wastewater (10kt)	X6
	Production of Industrial Waste Gas (100MB m^3)	X7
	Production of Industrial Solid Waste (10kt)	X8
State	Red Tide (time/year)	X9
	Cumulative Area of Sea Affected by Red Tide(m^2 /year)	X10
Response	General Water Quality in Sea Area (percentage of sea areas with water quality of worse than Grade IV and Grade IV)	Y

$$k_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \quad ...(1)$$

$$k_{ij} = \frac{x_{max} - x_{ij}}{x_{max} - x_{min}} \quad ...(2)$$

where, x_{ij} was a measured value of evaluation factor; x_{max} and x_{min} were the maximum and minimum measured value of evaluation factor, respectively; k_{ij} was the standardized index value.

Sample Setting

Training samples: Data-based machine learning is regarded as the research which looks for the rules from training data, and makes use of these rules for prediction. In order to ensure that the established SVM model is capable of fully utilizing the distribution feature of historical data and has the strong extrapolation capability, a part of historical data were converted into training samples by equation (3), to expand the training sample space. Training samples were used to perform pattern recognition on the SVM, so as to determine parameters and the parameters corresponding to kernel function.

$$\begin{cases} \Delta x_{i,j} = x_{i+j} - x_i \\ \Delta y_{i,j} = y_{i+j} - y_i \end{cases} i = 1,2,...,n; j = 2,3,...,n-i \quad ...(3)$$

As independent variable, x is an index value in the index system of marine ecological environment security prediction; as dependent variable, y is an integrated index value in this index system.

Test samples: The variation of adjacent samples in time series data, as the input/output, was substituted into the model for test, that is, equation (4) was used to get test samples. The model accuracy was tested according to the comparison between test value and true value.

$$\begin{cases} \Delta x_i = x_i - x_{i-1} \\ \Delta y_i = y_i - y_{i-1} \end{cases} i = 1,2,...,n \quad ...(4)$$

Simulated samples: Simulated samples were obtained through the conversion, equation (5), of each sample value in the influence factor matrix r. Taken as the output value after simulation for the trained model input, the simulated sample was converted to get the predicted ecological security value.

$$\Delta x_i' = x_{i+1} - x_i \quad ...(5)$$

Model recognition and model verification: The SVM's model recognition mainly determined the type and relevant parameters of kernel function type, and the selection of the

Table 2: Ningbo marine environment data from 2003 to 2013.

Year/Index	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Natural population growth rate	1.14	2.29	2.08	1.86	2.33	2.18	1.9	2.34	1.96	2.09	2.39
GDP per capita	32236	39046	43945	51285	61032	69997	73998	68162	77983	85475	93176
Total number of people	549.07	552.7	556.7	560.4	564.6	568.1	571	574.1	576.4	577.7	580.1
Sea area per capita	1499	1489	1479	1469	1458	1449	1442	1434	1428	1425	1419
Port throughput	18543	22586	26881	30969	34519	36185	38385	41217	43339	45303	49592
Total discharge amount of Industrial wastewater	11750	11528	11895	15827	17726	17290	17735	18970	19797	20125	19666
Production of industrial waste gas	2072	2046	2128	2678	3764	3753	4597	5349	5910	6218	6487
Production of industrial solid waste	1.14	2.29	2.08	1.86	2.33	2.18	1.9	2.34	1.96	2.09	2.39
Red tide	21	6	3	8	12	6	2	4	4	1	1
Cumulative area of sea affected by red tide	4026	2128	1618	2400	4962	3277	29	190	580	390	90
General water quality in sea area	84.56	60.62	58.52	50.03	52.23	30.01	58.80	71.33	56.41	80.4	44.41

parameters C and ε. Statistic standards of model parameter selection and model performance test included root-mean-square error (RMSE) and predicted residual sum of squares (PRESS), which are shown in equation (6) and equation (7). RMSE was used for model parameter selection, whereas PRESS was used for model performance test (Zhang et al. 2009).

$$PMSE = \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)2}{n-p}\right]^{\frac{1}{2}} \qquad ...(6)$$

$$PRESS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad ...(7)$$

where $\hat{y}_i$ is predicted value; y_i is true value; n is the number of samples; p is the number of model parameters.

***Selection of kernel function*:** Kernel function is the key for the SVM application and research, and directly determines final efficiency and performance. Most of the SVMs are nonlinear, and different kernel functions have their own features. Taking the research result by Yang et al. (2005) as the reference, we adopted the radial basis function (RBF) with strong nonlinear mapping capability and wide applications. Radial basis function:

$$K(x_i, x) = \exp\left(\frac{-\|x - x_i\|^2}{\sigma^2}\right) \qquad ...(8)$$

***Selection of parameters C and ε*:** The parameter is capable of controlling the penalty degree for samples exceeding the error ε. Theoretically speaking, if C value is too small, it will lead to the under-fitting of training data; if C value is too large, it will lead to over-fitting. Generally, the larger the ε value, the less the number of support vectors and the stronger the SVM's sparsity will be. However, if ε is too large, it will weaken the prediction performance of the SVM. There are many literatures reporting the method of selecting parameters (C, ε, r) for radius basis function, and grid search is the most common and reliable one. Grid search traverses all the possible combinations of (C, ε, r) to find out the parameters with the best performance. Parameter combination in this work was (100, 0.01, 1.4).

Reconstruction of multivariable time series and prediction on future attribute: Set multi-attribute time series sample unit as $R=\{R_1, R_2, ..., R_q\}$, index attribute set as $P=\{P_1, P_2, ..., P_m\}$, and time sample point as $T_i=\{1, 2, ..., l\}$. Marine ecological environment system is a complex nonlinear system. Every marine environment sample unit included m attributes for which the time series had a certain correlation. Therefore, the time series of m attributes of the same sample unit were converted into a m-dimensional multivariable time series for reconstruction so as to fill the gap of insufficient time series length. The embedding delay $\tau_j^{(k)}\,(j=1,2,...,m)$ and embedding dimension $d_j^{(k)}\,(j=1,2,...,m)$ of each attribute's time series $\left\{x_{ij}^{(k)}\right\}_{i=1}^{l}$ were evaluated respectively for multivariable delay reconstruction. The state vector in $d^{(k)}=\left(d_1^{(k)}+d_2^{(k)}+...d_m^{(k)}\right)$-dimensional phase space after the reconstruction was

$$X_n^{(k)} = \begin{pmatrix} x_{n1}^{(k)}, x_{n-\tau_1}^{(k)}, 1, ...x_{n-(d_1-1)\tau_1}^{(k)}, 1; \\ x_{n2}^{(k)}, x_{n-\tau_2}^{(k)}, 2, ...x_{n-(d_2-1)\tau_m}^{(k)}, 2; \\ .... \\ x_{nm}^{(k)}, x_{n-\tau_l}^{(k)}, m, ...x_{n-(d_m-1)\tau_l}^{(k)}, 2 \end{pmatrix}$$

$$n = \max_{1\le j\le m}\left(d_j^{(k)}-1\right)\tau_j^k + 1, ...l \qquad ...(9)$$

In the reconstructed phase space, with vector column $X_n^{(k)}$ as the input and $X_{n+1}^{(k)}$ as the output, SVM was used to estab-

Table 3: Comparison between true values and fitting results of SVM prediction model.

Year	General Water Quality in Sea Area /(percentage of sea areas with water quality of worse than Grade IV seawater and Grade IV) True Value	General Water Quality in Sea Area /(percentage of sea areas with water quality of worse than Grade IV seawater and Grade IV) Fitted Value	Fitting Error (%)
2003	0.85	0.84	2.90
2004	0.61	0.60	0.99
2005	0.59	0.58	1.20
2006	0.50	0.50	1.00
2007	0.52	0.52	0.96
2008	0.30	0.31	-1.02
2009	0.59	0.59	-0.85
2010	0.71	0.71	0.70
2011	0.56	0.57	-1.06
2012	0.80	0.61	24.63
2013	0.44	0.45	-0.90

Correlation Index R = 0.93; Determination Coefficient = 0.86

lish learning sample set $\left\{X_n^{(k)}, X_{n+1}^{(k)}\right\}, n = \max\limits_{1\le j\le m}\left(d_j^{(k)}-1\right)\tau_j^k+1,...l-1$. After determining kernel function and the parameters such as penalty factor C and loss coefficient e, sample training was performed. Then, using learning samples, the trained model constructed the relationship between $X_n^{(k)}$ and $X_{n+1}^{(k)}$, which met

$$x_{n+1\,j}^{(k)} = G_j\left(X_n^{(k)}\right) \qquad ...(10)$$

where $n = \max\limits_{1\le j\le m}\left(d_j^{(k)}-1\right)\tau_j^k+1,...l-1,\ j=1,2,...,m$. Finally, the trained SVM model (equation 10) can predict the value of each attribute of the sample unit $R_k\left(k=1,2,...q\right)$ in future time $l+1, l+2, l+3$.

ANALYSIS ON PREDICTION RESULTS OF SVM MODEL

The project performed model analysis on the data of ten factors related to Ningbo marine environment from 2003 to 2013 (Table 2), and calculated the percentage of sea areas with water quality of worse than Grade IV and Grade IV in Ningbo. The results are given in Table 3 and the change trend in prediction model is shown in Fig. 1. From Table 3, it was known that except for the year 2012, the average relative error of fitted values in prediction model was less than 3% in each year, with the correlation index of 0.93 and determination coefficient of 0.86. It meant a good fitting effect. As shown in Fig. 1, there was a high goodness of fit between the increased trend of Ningbo marine environment security prediction model and the increased trend of true values. Therefore, the SVM model can be used for Ningbo marine environment security prediction.

The SVM prediction model had a sound prediction performance for the characteristic quantity under multivariable influence. The multivariable SVM prediction model adopted by the project had high accuracy in predicting the characteristic quantity under multivariable influence in the marine environment (Wang et al. 2012).

According to the established Ningbo marine environment security prediction model, the percentage of seawater of worse than Grade IV and Grade IV in Ningbo sea area from 2014 to 2016 was predicted, with the results presented in Table 4. It can be seen that the percentage of seawater of worse than Grade IV and Grade IV from 2014 to 2016 presented an increase trend. The related departments should properly treat the discharged industrial wastewater, and the produced industrial waste gas and solid waste. Otherwise, these pollutants will threaten the Ningbo marine ecological environment security.

DISCUSSION AND CONCLUSION

(1) Using the SVM model, learning, simulation and prediction were performed on the development trend of each system of Ningbo marine ecological security. The result showed that the relative error of the SVM prediction model was basically less than 3% with the correlation index of 0.93 and determination coefficient of 0.86, indicating a high goodness of fit. The prediction result accorded with the actual situations in Ningbo and was basically capable of reflecting the overall development trend of Ningbo ecological security. This confirms that SVM is an effective method of marine ecological security prediction.

(2) From 2003 to 2013, the percentage of seawater of worse than Grade IV and Grade IV in Ningbo sea area decreased slightly on the whole, but the reduction speed was limited. From 2014 to 2016, this percentage was about 50%, which means that Ningbo marine ecological environment quality was not optimistic.

(3) In the recent 12 years, the development trend of Ningbo

Table 4: Prediction on the general water quality in Ningbo sea area from 2014 to 2016 (percentage of seawater of worse than Grade IV and Grade IV).

Year	Index: General water quality in sea area (percentage of seawater of worse than Grade IV and Grade IV)
2014	46.01
2015	49.74
2016	52.72

ecological security has not been fundamentally changed, which has a close relationship with the factors such as rapid industrial development, large discharge amount of industrial waste gas and wastewater, and fast population increase. Therefore, the counter measures should be positively adopted to relieve Ningbo marine ecological pressure.

Problems in the Project

(1) The SVM method establishes the prediction model by continuously learning historical samples, and the establishment of the SVM prediction model is a process of self-learning and memory of historical samples. A larger number of training samples will make the prediction model cover more support information so as to improve the prediction accuracy. Therefore, new samples should be continuously added for training in real predictions, thus continuously improving the prediction model.

(2) Although the SVM method has no definite limitation on factor quantity, it does not mean that more factors will bring about a better result. It has been found that the introduction of the factors with weak resolution capability has the possibility of adding noises and hence influences the prediction effect. If factor analysis is performed in advance to reject some factors, the prediction effect will be more reasonable.

(3) Comprehensive and effective evaluation index system is the key to ensure that prediction result can conform to the reality. However, influence by many elements, some data were hard to collect, and some indices were not considered comprehensively, so the index system established may not be perfect. Further improvement in future is required. Meanwhile, this article only predicted and analyzed the development trend of Ningbo marine ecological security, and there is the need to further discuss the problems such as driving force and regulatory mechanism of marine ecological security in future. We only discussed the ecological security of Ningbo marine resources, but the other relevant security problems also deserved to be explored systematically.

ACKNOWLEDGEMENTS

This paper belongs to the project of the “Zhejiang public welfare funding project (No. 2014C31161).

REFERENCES

Chen, Xiuzhong 2012. Learning the spirit of 18th CPC national congress to help Ningbo marine economy development. China Ocean News, pp. 1.

Dai, Chao, Liu, Hui and Yang, Shu-qin 2013. The status of Ningbo’s environmental pollution and its counterplan research. Journal of Jiangsu Teachers University of Technology, 3(19): 57-63.

Coetes, C. and Vapnik, V. 1995. Support vector networks. Machine Learning, 20(3): 273- 297.

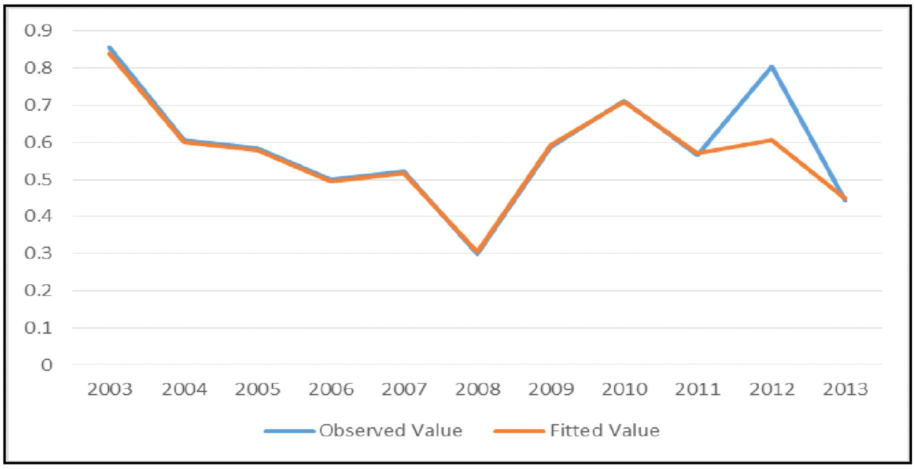

Fig. 1: Comparison between the change trends of true and fitted values of the percentage of seawater of worse than Grade IV and Grade IV.

Raudys, S. 2000. How good are support vector machines. Neural Networks, 13(1): 17- 19.

Burges, C. J. C. A. 1998. Tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2): 121-167.

Ma, Zhi-jiang, Chen, Han-lin, and Yang, Shu-feng 2003. Prediction of landslide hazard based on support vector machine theory. Journal of Zhejiang University (Science Edition), 30(5): 592-596.

Zhang, Rui, Zheng, Hua-wei, and Liu You-zhao 2014. Health evaluation of land use system based on pressure-state-response model and set pair analysis. Bulletin of Soil and Water Conservation, 34(5): 146-152.

Zhang, Zhi-yong and Liu, Xi-yu 2009. Research on land resource eco-security assessment based on support vector machines. Computer Engineering and Applications, 45(10): 245-245.

Ling, Wu-neng, Hang, Nai-shan and Li, Ru-qi 2013. Short period wind power prediction based on cloud support vector machine. Electric Power Automation Equipment, 33(7): 34-38.

Yang, Yi-wen and Yang, Zhao-jun 2005. Financial time series forecasting based on support vector machine. System Engineering - Theory, Methodology, Applications, 14(2): 177-181.

Wang, Peng and Li, Zhong-hua 2012. Research and application of prediction ability of multivariable support vector machine. Jiangsu Science and Technology Information, 8: 37-40.