Prediction Model of Alga's Growth Based on Support Vector Regression

Yan Qisheng
School of Mathematics & Information Science
East China Institute of Technology
Fuzhou 344000, China
School of Information
Jiangnan University, Wuxi 214122, China
e-mail: yanqs93@126.com

Wang Guohua
School of Mathematics & Information Science
East China Institute of Technology
Fuzhou 344000, China

***Abstract*—Support vector machines are a kind of novel machine learning methods, which have become the hotspot of machine learning because of their excellent learning performance. A alga's growth prediction model is established by using support vector regression (SVR) method. The method is illustrated through examples, the results obtained by using SVR method are compared with that from neural network method and the results show that the prediction model based on support vector regression is more accurate and simple than neural network method.**

Keywords- support vector regression;prediction model;alga's growth;neural network

I. INTRODUCTION

Nowadays, artificial neural networks are being extensively used in many fields because of their massively parallel processing capability and approximation ability for arbitrary function and the characteristics of self-learning and adaptive. However, neural networks as a kind of heuristic technique relayed on experience face on some problems such as over-fitting phenomena and resulting in bad generalization ability at present [1]. Support vector machines proposed by Vapnik[2] are a kind of novel machine learning methods based on statistical theory which have become the hotspot of machine learning because of their excellent learning performance. In recent years there has been an increasing interest in studying support vector machines(SVM) for classification and regression problems and numerous applications have been reported in various fields of science and engineering [3,6]. It transforms the optimization problem into a convex quadratic programming problem and the global optimal solution can be uniquely obtained. So support vector machines can avoid over-fitting and have better generalization ability as compared with neural networks. In this paper, support vector regression method has been tried to apply to construct prediction model of alga's growth based on small sample.

II. SUPPORT VECTOR REGRESSION MODEL

A. Linear regression problems

Consider the problem of approximating the set of data,

$$D=\{(x_i,y_i)\mid i=1,2,\cdots,n\}\ ,\quad x_i\in R^n,\ y_i\in R\ ,$$

with a linear function,

$$f(x)=\langle \omega,x\rangle+b \tag{1}$$

SVM can be applied to regression problems by the introduction of an alternative loss function. The loss function must be modified to include a distance measure. Four possible loss functions are ε — insensitive loss function, Quadratic loss function, Huber loss function and Laplace loss function etc [4].

Using an ε —insensitive loss function,

$$L_\varepsilon(y)=\begin{cases}0 & ,\ \ if\ \ |f(x)-y|<\varepsilon\\ \big||f(x)-y|-\varepsilon\ , & otherwise\end{cases} \tag{2}$$

the optimal regression function is given by the minimum of functional,

$$\min_{\omega,b,\xi_i,\xi_i^*}\Phi=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$$

with constraints,

$$\begin{aligned}&((\omega\bullet x_i)+b)-y_i\le\varepsilon+\xi_i\ , && i=1,2,\cdots,n\\ &y_i-((\omega\bullet x_i)+b)\le\varepsilon+\xi_i^*\ , && i=1,2,\cdots,n\\ &\xi_i,\xi_i^*\ge 0\ , && i=1,2,\cdots,n\end{aligned} \tag{3}$$

or alternatively,

$$\max_{\alpha,\alpha^*}W=-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle+\sum_{i=1}^{n}[\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)] \tag{4}$$

with constraints

978-0-7695-3682-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ESIAT.2009.170

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0,\ 0 \le \alpha_i, \alpha_i^* \le C,\ \ i = 1,2,\cdots,n$$

(5)

Solving Equation (4) with constraints Equation (5) determines the Lagrange multipliers, α_i, α_i^* , and the regression function is given by Equation (1), where

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i$$

(6)

The Karush-Kuhn-Tucker (KKT) conditions that are satisfied by the solution are,

$$\alpha_i \alpha_i^* = 0, \qquad i = 1,2,\cdots,n$$

(7)

Therefore the support vectors are points where exactly one of the Lagrange multipliers is greater than zero.

When ε=0, we get the loss function and the optimization problem is simplified,

$$\min_{\beta} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\left\langle x_i, x_j \right\rangle - \sum_{i=1}^{n}\beta_i y_i$$

(8)

with constraints, $-C \le \beta_i \le C, \qquad i = 1,2,\cdots,n$

(9)

$$\sum_{i=1}^{n}\beta_i = 0,$$

and the regression function is given by Equation (1), where

$$\omega = \sum_{i=1}^{n}\beta_i x_i,\quad b = -\frac{1}{2}\left\langle \omega, (x_r + x_s) \right\rangle$$

(10)

B. Non linear regression problems

For non linear regression, regression equation,

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

(11)

The non-linear SVR solution, using an ε —insensitive loss function, is given by,

$$\max_{\alpha,\alpha^*} W = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + \sum_{i=1}^{n}[\alpha_i(y_i - \varepsilon) - \alpha_i^*(y_i + \varepsilon)]$$

with constraints,

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0,\ 0 \le \alpha_i, \alpha_i^* \le C,\ \ i = 1,2,\cdots,n$$

(12)

where $K(x_i, x)$ is a kernel function. The frequently-used kernel functions include:

1) Linear kernel function: $K(x, y) = xy$

2) Polynomial kernel function:

$$K(x, y) = (xy + 1)^d, \quad d = 1,2,\cdots$$

3) Radial Basis kernel function:

$$K(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2})$$

4) Multi-Layer Perceptron kernel function and Splines kernel function etc.

III. APPLY SVR TO THE PREDICTION OF ALGA'S GROWTH

A. Prediction of Alga's Growth Based on BP Neural Network

Based on artificial neural network theory [1], the paper proposes a growth prediction model of alga in the Yellow River. TABLE I illustrates the numbers (10e6 ind /L) of alga in the Yellow River from February to October during 2000-2003. In this paper, the numbers (10e6 ind /L) of alga for three months (January, November and December) are not considered because the alga is scarce in winter. The adopted neural network architecture (with two hidden layer) is 3-3-3-1, namely there are three neurons in input layer, each hidden layer includes three neurons and output layer includes one neuron. In this section, we will use the data of each month to train neural network independently because the data has larger difference. Specifically speaking, the data of some month of the previous three years (input data) and the data of the corresponding month of fourth years (output data) have been used to train neural network [5]. The parameter setting is: error_goal = 0.05, lr = 0.05, max_epoch = 50. The simulation results using Matlab were given in TABLE II.

B. Prediction of Alga's Growth Based on SVR

The same data as above Table.1 are used to simulate. In this Section, we take the data of the same month during 2000-2002 as the input of SVR and take the data of the corresponding month as the desired output of SVR, and then the simulation study is carried out. where $n = 9$, $K(x_i, x)$ using Radial Basis kernel function which form is given by $K(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2})$, parameter C is selected as 1000, $\varepsilon = 0.01, \sigma = 1$. TABLE II shows the simulation results using Matlab based on SVM and BP neural network respectively.

From the simulation results we can see that the SVR method has more exact result than BP neural network method. So these results suggest that SVR method has a better ability of generalization. It is easy to predict alga growth situation after 2003 using SVR method.

IV. CONCLUSIONS

In this paper, a alga's growth prediction model based on support vector regression was presented aiming at neural network affected by network structure and the complexity of sampling, especially small sample statistical learning be easy to recur over fitting and low generalization. The experimental result demonstrates that SVM method is more exact than BP neural network for the prediction of alga's growth. However, just as the other algorithm to solve the pattern recognition problems, SVM method faces on the problem of parameter selection and optimization [6].

REFERENCES

[1] Simon Haykin. Neural Networks: A Comprehensive Foundation, Second Edition. Beijing: China Machine Press, 2004

[2] Vapnik V N. The Nature of Satistical Learning Theory. New York Springer Verlag, 1995

[3] Martinez, D. and A. Bray. "Nonlinear blind source separation using kernels". IEEE Transactions on Neural networks, 2003, 14(1), pp.228-235

[4] Gunn S R. "Support Vector Machines for Classification and Regression". Technical Report, Image Speech and Intelligent Systems Research Group, university of Southampton, 1997

[5] YANG Yin- dong, ZHOU Yong- quan, LIU Yong- bin. "An Improved Method based on BP Neural Network for Forecasting Alga's Growth". Modern Computer, 2006.9, pp.11-13(in Chinese)

[6] N.E. Ayat, M. Cheriet, C.Y. Suen, "Automatic model selection for the optimization of SVM kernels". Pattern Recognition, 2005, 38 (10), pp.1733−1745

TABLE I. THE NUMBERS (10E6 IND /L) OF ALGA FROM FEBRUARY TO OCTOBER DURING 2000-2003

year	2	3	4	5	6	7	8	9	10
2000	12.5	110	4.6	6.1	8.8	24.7	76.2	166	9.1
2001	13.1	112	5.3	6.7	9.1	25.4	77.6	169	9.4
2002	13.4	113	5.8	7.1	9.8	26.2	78.1	174	9.5
2003	13.7	123.3	5.5	7.3	9.9	26.7	84.3	169	9.0

TABLE II. TABLE.2 COMPARISON OF PREDICTION RESULTS BASED ON SVR AND BP

month	The year 2000	The year 2001	The year 2002	The year 2003 actual value	SVR method		Bp neural network method	
					predictive value for the year 2003	Relative error	predictive value for the year 2003	Relative error
2	12.5	13.1	13.4	13.7	13.69	-0.073%	13.5063	-1.414%
3	110	112	113	123.3	123.29	-0.08%	123.0883	-0.172%
4	4.6	5.3	5.8	5.5	5.49	-0.182%	5.3123	-3.413%
5	6.1	6.7	7.1	7.3	7.29	-0.137%	7.1169	-2.508%
6	8.8	9.1	9.8	9.9	9.89	-0.101%	9.7177	-1.841%
7	24.7	25.4	26.2	26.7	26.69	-0.37%	26.4807	-0.821%
8	76.2	77.6	78.1	84.3	84.29	-0.012%	84.1100	-0.225%
9	166	169	174	169	168.99	-0.006%	168.8077	-0.114%
10	9.1	9.4	9.5	9.0	8.99	-0.111%	8.8082	-2.131%