Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 31 (2012) 1194 – 1199

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Advances in Computational Modeling and Simulation

Prediction of water quality time series data based on least squares support vector machine

Guohua Tan*, Jianzhuo Yan, Chen Gao, Suhua Yang

College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing ,China

Abstract

Actual water quality monitoring sites due to inadequate and abnormal by the impact to water quality warning has brought new challenges. How do based on limited monitoring data to predict the water quality to address the shortage of river water quality monitoring sites and data coverage of false alarms caused by abnormal, early warning of water pollution are of great significance. According to river water monitoring data, the small sample properties, is proposed based on least squares support vector machine prediction of water quality parameters, this method has strong ability to predict the true value, and the global optimization and good generalization. This method is applied in the river water quality measurement data, after training the LS-SVM model for water quality parameters of water quality monitoring system to predict, in the same sample under the BP network and RBF network prediction. Experimental results show that the small sample case with noise, least squares support vector machine method is better than multi-layer BP and RBF neural network, to better meets the requirements of water quality prediction. Comparing data of the two experiments shows that new module makes the interest mining more effective.

Keywords: Support Vector Machines, Time series data forecast, Small sample, Water Quality Monitoring,

*Corresponding author. Tel:+86-15210835051
Email address: tgh@emails.bjut.edu.cn

doi:10.1016/j.proeng.2012.01.1162

1. Introduction

Forecasting changes in water quality of water environment, planning, evaluation and management of the foundation is to maintain and manage the current basis for water quality, water quality by predicting the evolution of trends can be understood in order to find the reasons for deterioration of water quality, and can guide the development of control measures.

But in the actual water quality control, due to the limited number of water quality monitoring sites, resulting in lack of monitoring data, monitoring data also lead to abnormal environmental impact, which bring more to the prediction of water quality challenges.

The current domestic small samples there are two categories of prediction methods:

(1)Prediction model based on neural networks[1~2], BP networks using nonlinear function approximation theory to predict the sample.

(2)Using principal component analysis (PCA) to model the main features of extracted information, high dimensional complex data dimension reduction, data reconstruction through the correct estimate of the data.

In recent years, researchers at home and abroad to support vector machine classification and regression were a large number of studies, such as literature [3~4] to support vector machines, artificial neural networks have been studied, indicating that the support vector machine classification and regression have a greater advantages, can solve the small sample, nonlinear, high dimension and local minimum points and other practical issues, is considered a better alternative to artificial neural network algorithm.

In this paper, least squares support vector machine LS-SVM algorithm to construct non-linear time series forecasting model to predict the water quality.

2. Forecasting methods and implementation process

2.1 Phase space reconstruction theory

Phase space reconstruction theory [5] is the basis for time series prediction, Packard and others and Takens[6] delay coordinates proposed method using time series phase space reconstruction. Phase space reconstruction theory is that the system is the evolution of any component interact with other components of the decision, therefore, the evolution of each component in the system are implicit in all of the information. When a state space reconstruction, only consider a component, and it is a fixed time delay in some point as a new dimension processing.

In this paper, in order to reduce modeling errors, the first for the raw data processing and data zero-mean normalization, and then according to Takens theory of phase space reconstruction, about one-dimensional time series into a matrix form to obtain the relationship between data relations, and thus tap into as much as possible the amount of information.

2.2 LS-SVM time series prediction algorithm

Suyken et proposed least squares support vector machines[7~9] in recent years, important results of statistical learning theory is one, least squares support vector machine training process follows the structural risk minimization principle, and the general vector machine has the computational complexity compared to low, the advantages of high operation speed.

Least squares support vector machines time series prediction algorithm described as follows:

Training samples for linear regression problems, training data set $(x_i, y_i), i = 1, 2 \ldots n, x_i \in R^d$ is the i sample of the input mode, $y_i \in R$ corresponding to the desired output samples.

Linear regression function:

$$y(x) = w^T x + b \quad [1]$$

According to SRM criteria will be returned to the problem into the following constrained quadratic optimization problem, it is the only optimal solution exists:

$$\min J(w, \xi) = \frac{1}{2} w^t w + \frac{\gamma}{2} \sum_{i=1}^{n} \xi_i^2 \quad [2]$$

Constraints:

$$yi = w^T x_i + b + \xi_i, i = 1, \ldots, n \quad [3]$$

LS-SVM regression, with a low-dimensional space to high dimensional space (Hilbert space) [10]of the nonlinear mapping $\varphi(\cdot) : R^d \to R^{dh}$, the low-dimensional nonlinear regression into high-dimensional space, linear regression, the definition of high-dimensional space for the inner product operation $K(x_i, x_j) = (\varphi(x_i)\varphi(x_j))$, where (x_i^T, x_j) replacing the inner product operation $K(x_i, x_j)$, the introduction of Lagrange function, the constrained optimization problem into the linear equation obtained nonlinear LS-SVM regression model:

$$f(x) = \sum_{i=1}^{n} a_i K(x_i, x) + b \quad [4]$$

Type in the Mercer kernel function is to satisfy the conditions can be any symmetric function, we use the radial basis function RBF:

$$K(x_i, x_j) = \exp\left\{ -\frac{\left| x_i, x_j \right|^2}{2\sigma^2} \right\} \quad [5]$$

2.3 Implementation process

Based on the above algorithm, this paper proposes using least squares support vector machine LS-SVM (Least Squares Support Vector Machine) algorithm to construct non-linear time series forecasting model to predict the water quality. The method mainly consists of the following modules completed, the overall implementation framework shown in Figure 1.

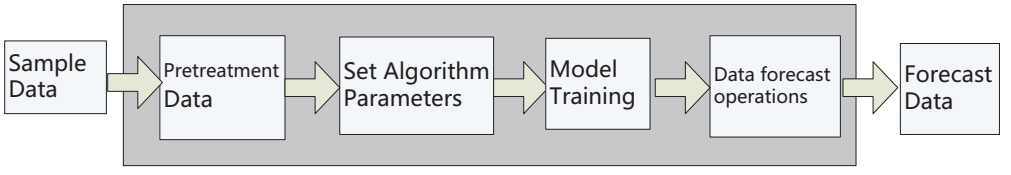

Fig. 1 Implementation framework of predict the water quality

The above frame mainly has four steps, Pretreatment Data, Set Algorithm Parameters, Model Training and Data forecast operations. The details are below.

2.3.1 Pretreatment Data

Construction of training samples under the conditions of small samples, to validate the model and the real value of the predicted anti-noise performance, the training samples and test samples to make the

following treatment:

(1)The experimental data were first pre-entered; select the data and the normalized, after transformation, the data values in between 0 and 1.

(2)In the measurement data, the variance of 0.01 were added random white noise, the direct use of these data with the noise form the training samples

2.3.2 Set Algorithm Parameters

LS-SVM model with the measured data to predict water quality, the use of radial basis function as the RBF kernel function, the use of LSSVMLab matlab toolbox algorithm framework as the basic package, parameter Gam=10, sig2=0.5, Type='function estimation'; with BP Neural network model forecasts of water quality measurements, using a standard3-layer BP network, an input node, an output node, 20 hidden nodes; with RBF neural network model forecasts of water quality measurement data, using generalized regression neural Network GRNN (Generalized Regression NN).

2.3.3 Model Training

Using trainlssvm function training the network, the regression coefficient and deviation b, so as to arrive as in [4] shows the prediction model.

2.3.4 Data forecast operations

Trained by the previous LS-SVM, BP and RBF models predict water quality parameters of the experimental data.

3. Experimental Results and Analysis

The LS-SVM algorithm is applied to a section of the river water quality monitoring data to predict the total phosphorus, and with the BP (Back Propagation), RBF (Radial Basis Function) network method, a comparative study of prediction.

3.1 Experimental Results

LS-SVM, BP network and RBF network training for the three models and error learning curve shown in Figure 2, "Solid line" represents the total phosphorus in water quality monitoring parameters of the actual value, "Dotted line" represents the predicted value, its deviation from the reflection of the phosphorus measured deviation from the actual value of the degree of the graph shows the deviation from the actual deviation value has no strict counterpart, because after the data preprocessing and phase space reconstruction after the great changes of their range; eight output of the test sample and the relative error of the forecast shown in Table 2 are 0.064 percent, 0.529 percent and 0.552 percent,experimental results Table 2 below.

The total phosphorus in water quality monitoring parameters of the actual value, see Table 1.

Forecasting value comparison of LS-SVM,RBF and BP network, see Table 2.

Table 1 The total phosphorus in water quality monitoring parameters of the actual value Model.

Month	1	2	3	4	5	6	7	8	9	10

TP	0.0763	0.0312	0.0607	0.0380	0.1260	0.1152	0.1280	0.1172	0.1273	0.1233

Input of the three model simulation compared with the predicted value of the map shown in Figure 2.

The first chart is LS-SVM model, the second chart is BP network, the third chart is RBF network.

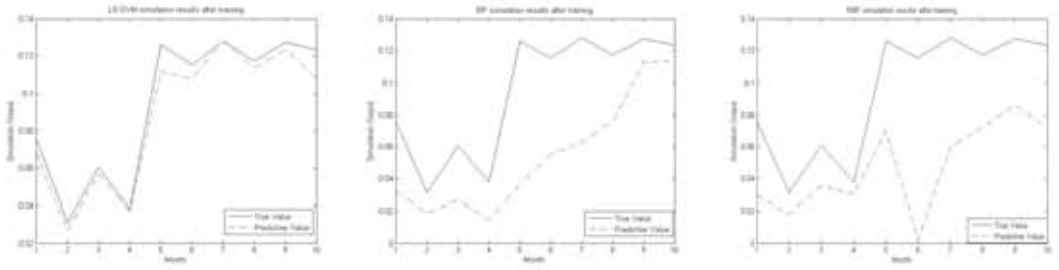

Fig. 2 models the input and output value of the comparison chart

The output of the model are compared with the input value in three models obtained the best predictive model output compared in Table 2.

Table 2 Forecasting value comparison of LS-SVM,RBF and BP network

Model		LS-SVM		BP network		RBF network	
Month	True Value/ (mg/L)	Predictive Value/ (mg/L)	Relative Value/%	Predictive value/ (mg/L)	Relative Value/%	Predictive value/ (mg/L)	Relative Value/%
1	0.0763	0.0686	0.10	0.0320	1.42	0.0297	1.39
2	0.0312	0.0268	0.14	0.0184	0.41	0.0180	0.42
3	0.0607	0.0570	0.06	0.0267	0.56	0.0358	0.41
4	0.0380	0.0372	0.02	0.0144	0.62	0.0307	0.19
5	0.1260	0.1121	0.11	0.0365	0.71	0.0705	0.44
6	0.1152	0.1082	0.06	0.0552	0.52	0.0023	1.02
7	0.1280	0.1280	0.00	0.0627	0.51	0.0601	0.53
8	0.1172	0.1136	0.03	0.0761	0.35	0.0726	0.38
9	0.1273	0.1234	0.03	0.1132	0.11	0.0865	0.32
10	0.1233	0.1085	0.12	0.1134	0.08	0.0715	0.42
Average Value	—	—	0.064	—	0.529	—	0.552

3.2. Comparative Analysis

Comparison of experiment with LS-SVM can be seen to predict the time series based on the true value of water quality measurement data, the data fit a good degree of good, better than the BP network method and the RBF Network. This is mainly because the LS-SVM model can take full advantage of the distribution of training samples, based on discriminate function constituted part of the training samples, and the LS-SVM model processing ability of small samples, in the case of small samples can be well Conduct statistical learning.

4. Conclusions

Based on the timing of water quality monitoring data with small samples and noise characteristics of

the data presented method based on LS-SVM time series prediction model for water quality monitoring. It will be by means of phase space reconstruction Reconstruction of time series data for the vector data, obtained through the normal LS-SVM model to learn on small samples, and can produce good performance of the prediction model, simulation results show that the LS-SVM based Water quality prediction model, root mean square error and mean relative error than the BP network method, RBF network method is much smaller, indicating that the LS-SVM method has a high prediction accuracy, more applicable to real-time water quality data with small sample forecast.

Acknowledgements

The data of Beijing Water Authority, and the matlab experiment framework used LSSVMLab matlab toolbox download from http://www.esat.kuleuven.be/sista/lssvmlab/,a special thanks here.

References

[1] Peng T M， Hubele N F,Karady G G.Advancement in the application of neural networks for short-term load forecasting.IEEE Trans on Power Systems,1992,7(1):250～257.

[2] Park D C,El-Sharkawi M A et al.Electric load forecasting using an artificial neural networks.IEEE Trans on Power Systems,1991,6(2)：442～448.

[3] J Tian Han，13o—Suk Yang，Jong Moon Lee．A new condition monitoring and fault diagnosis system of induction motors using artificial intelligence algorithms[J]．Electric Machines and Drives，2005：1967—1974

[4] Ze Dong，Pu Elan，Xi—Chao Yin．Simulation study on sensor fault diagnoses of the temperature of the boiler high—temperature part metal wall[J]．Machine Learning and Cybernetics，2004，5 (26— 29)：3003—3008.

[5] Takens F. Detecting Strange Attractors in Turbulence [C]// Dynamical Systems and Turbulence, Lecture Notes in Mathematics. Berlin: Springer-Verlag, 1981, 898: 366-381.

[6] Packard N H, Crutchfield J P, Farmer J D, et al. Geometry From a Time Series [J]. Physical Review Letters (S0031-9007), 1980, 45(9): 712-716.

[7] Vapnik V N．An overview of statistical learning theory[J].IEEE Trans. on Neural Networks,1999,10(5):988-999.

[8] Platt J.Fast training of support vector machine using sequential minimum optimization[C]．Advance in Kernel Methods-support Vector Learning, Cambridge, 1999:185-208.

[9] Suykens J A K,Lukas L,Vandewalle J.Sparse approximation using least squares support vector machine[C].IEEE Int. Symposium on Circuit and Systems(ISCAS 2000),Geneva,Switzerland,2000.

[10] R. J. Duffin and A. C. Schaeffer, A class of no harmonic Fourier series, Trans. Amer. Math. Soc. 72 (1952), 341 - 366.