Sensors & Transducers, Vol. 172, Issue 6, June 2014, pp. 290-294

Sensors & Transducers

An Artificial Neural Network Based Short-term Dynamic Prediction of Algae Bloom

Yao Junyang, * Shi Yan, Wang Xiaoyi, Xu Jiping
Beijing Technology & Business University, College of Computer and Information Engineering,
Beijing, 100048, China
Tel.: 13691215263
* E-mail: shiyanspring@126.com

Received: 11 April 2014 /Accepted: 30 May 2014 /Published: 30 June 2014

Abstract: This paper proposes a method of short-term prediction of algae bloom based on artificial neural network. Firstly, principal component analysis is applied to water environmental factors in algae bloom raceway ponds to get main factors that influence the formation of algae blooms. Then, a model of short-term dynamic prediction based on neural network is built with the current chlorophyll_a values as input and the chlorophyll_a values in the next moment as output to realize short-term dynamic prediction of algae bloom. Simulation results show that the model can realize short-term prediction of algae bloom effectively. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Algae bloom prediction, Principal component analysis (PCA), Radial basis function neural (RBF) network.

1. Introduction

A lot of nitrogen, phosphorus and other nutrients are discharged into water with the economic and social development, which causes algae and other plankton proliferate and leads to increasingly serious eutrophication degree and frequent algae blooms. The development of city construction and people's normal life are directly influenced because of the threats to water ecological security [1]. Therefore, the study of the formation of algae bloom and effective prediction of algae bloom, the unconventional emergency, is of great significance. Due to various environmental factors and complex role relationship that influence the growth of algae bloom, it is often difficult to obtain satisfactory prediction accuracy based on the growth mechanism of algae bloom. Prediction method based on data driven, by contrast, is not dependent on the growth mechanism of the algae bloom and to establish the optimal mathematical relationship between the input and output data as the goal, is applicable to prediction problem. Ren et al established BP network prediction model of algae bloom for The Three Gorges [2]. The model predicted the trend of eutrophication of The Three Gorges Reservoir in 2005 with eight system parameters including dissolved phosphorus, colloidal phosphorus, total nitrogen (TN), chemical oxygen demand (COD), flow velocity, water depth and temperature as the input variables, and index representing the degree of eutrophication of water body as output. Pei et al established neural network model and predict the concentration of chlorophyll_a in West Lake by repeated training to find water quality factors that reflects the trends of aquatic ecological status [3]. Li et al established artificial network prediction model of algae bloom with temperature, dissolved oxygen, nitrogen-phosphorus ratio, total nitrogen and light as the input variables, and chlorophyll_a as index for the characterization of blooms [4].

But the prediction models of algae bloom above considered less affecting factors and could not fully reflect water features and the trend of blooms outbreak. This paper uses the data from raceway experiments which is released by NASA's website. The data includes 13 parameters and the algae bloom was cultivated by the American Type Culture Collection. Principal Components Analysis is applied to extract the main water environmental information and RBF neural network is built to predict the value of chlorophyll_a. This paper can finally realize short-term dynamic prediction of algae blooms.

2. The Determination of Main Factors that Influent the Formation of Algae Blooms

2.1. Obtaining of Forming Indices of Blooms

Algae bloom was cultivated respectively in three raceway ponds in the same conditions and three sets of data is obtained, recorded as algae, algae_kg and algae_ts. There are thirteen parameters including irradiance, temperature, pH, oxygen, dissolved inorganic carbon (DIC), nitrogen, phosphorus, density, salinity, surface measurement of pulse amplitude modulation (PAM0), middle measurement of pulse amplitude modulation (PAM8), bottom measurement of pulse amplitude modulation (PAM16) and chlorophyll_a in the data. Chlorophyll_a is the most direct characterization of the existing amount of algae in water, so it can be used as the characteristic index of judging the formation of algae blooms [5]. The data was collected with different timing. Chlorophyll_a was measured in days, so the sampling intervals of other indicators are unified to process to measure in days. Temperature and irradiance sensors took measures every 5 minutes, while all the other data were collected, most times, 3 times per week at the same time.

The cyanobacterium S. platensis 29408 was obtained from ATCC. This strain was isolated from a saline marsh, in DelMar Slough, CA. Three raceways were constructed and operated in a rooftop greenhouse. The raceway ponds have a 200 liter capacity and a paddle wheel with four blades that rotates at a speed of 6.785 rpm. All raceways were inoculated with 1.5 liters of S. platensis culture and 200 L of S. platensis media. Automated and manual measurements were performed from the three raceways throughout the S. platensis harvesting cycle. Automated measurements of for temperature and irradiance were downloaded from a datalogger weekly. Manual measurements of pH, salinity, oxygen concentration, nutrients, dissolved inorganic carbons, chlorophyll_a, and optical density were performed every day at the same time. Air temperature was the only controlled factor throughout the duration of the experiment.

2.2. Obtaining of Main Factors that Influence the Formation of Algae Blooms

Principal Component Analysis (PCA) aims to use the idea of reducing dimension to convert the multi-index into a few more comprehensive indicators, which fundamentally solved the problem of the information overlap between the indicators and greatly simplified the structure of the original index system [6-7].

This paper uses Principal Component Analysis, with irradiance, temperature, pH, oxygen, DIC, nitrogen, phosphorus, density, salinity, PAM0, PAM8, PAM16 as original variables for data algae, to convert the multi-variables into a few more comprehensive variables, which means that water environmental information dimensions is reduced and the ingredients playing a leading role from a number of factors in the growth of algae blooms are identified.

When the data algae is applied to principal component analysis, the eigenvalues of first three principal components accounts for more than 95 % of the total variance, as shown in Table 1. In other words, the first three principal components cover most of the pollutants information of the twelve monitoring indicators. So the first three principal components are chosen as the main evaluation index. Principal component analysis results are shown in Table 2.

Table 1. Principal component loading matrix.

Principal components	Eigenvalue	Contribution rate (%)	Total variation (%)
1	9837.372	0.7062	0.816137
2	2216.214	0.1591	0.865242
3	1183.479	0.0850	0.950195
4	648.6617	0.04656	0.996758
5	27.60234	0.00199	0.998739
6	7.85647	0.0006	0.999303
7	7.381014	0.0005	0.999833
8	2.208357	0.0002	0.999992
9	0.08453	6.0678E-06	0.999998
10	0.02203	1.5814E-06	0.999999
11	0.005823	4.1796E-07	1
12	0.00169	1.2129E-07	1

Table 2. Principal component loading matrix.

Index	The first principal component	The second principal component	The third principal component
Irradiance	0.0031	0.4644	0.0475
Temperature	5.7407e-05	0.0105	6.3181e-05
pH	8.4035e-06	0.0012	5.3153e-05
Oxygen	0.0059	0.75267	0.5046
DIC	0.9999	-0.0079	0.0005
Nitrogen	0.0009	0.0082	0.0049
Phosphorus	4.8398e-05	0.0011	0.0008
Density	0.0001	0.0020	0.0008
Salinity	0.0002	0.0009	0.0010
PAM0	0.0041	0.4466	0.8219
PAM8	0.0005	0.0536	0.0829
PAM16	0.0010	0.12346	0.2464

3. Artificial Neural Network Based Short-term Algae Bloom Prediction

3.1. The Construction of Artificial Neural Network Based Short-term Algae Bloom Prediction

Radial Basis Function (RBF) neural network is a kind of feed forward neural network with topology. It is composed of input layer, hidden layer and output layer. Only a small amount of weight need to be adjusted for each input and output data, and the neural network has the advantage of fast learning speed, global approximation and best approximation performance, at the same time, the training method is rapid and feasible [8-10].

This paper takes RBF neural network to build prediction model of algae bloom, and the network structure is shown in Fig. 1.

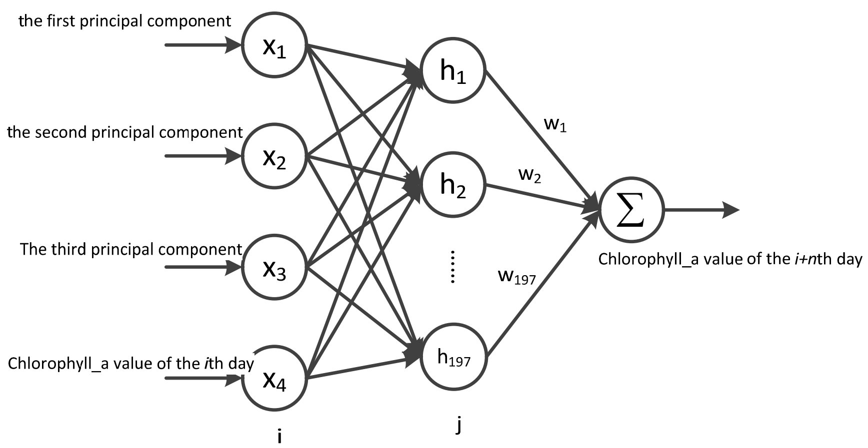

Fig. 1. RBF neural network structure.

This neural network takes the value of principal components and chlorophyll_a in ith day as input and the value of chlorophyll_a in i+nth day as output, in which the mean square error is set zero, expansion rate of radial basis function is 250, maximum number of neurons is 197, and the number of neurons between the two displays is 25.

3.2. Simulation

The observation time of water quality parameters of the raceway ponds was from January 15, 2010 to October 15, 2010, totally 273 days. This paper focuses on the outbreak of algae bloom prediction, so the data of first 200 days including lag or induction phase, exponential phase, phase of declining relative growth and stationary phase is selected. The selected data is used to predict the outbreak of algae bloom as training sample. This paper uses the value of principal components and chlorophyll_a in ith day of data algae as input to train the network, and then takes data algae_kg as test samples to predict the value of chlorophyll_a in i+3th day. The flow chart of the process is shown in Fig. 2. In order to verify the validity of the model, at the same time, the results are compared with BP network's, and the result is shown in Fig 3.

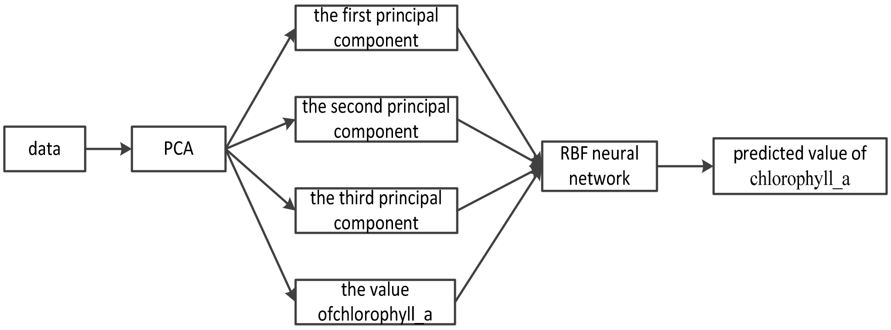

Fig. 2. The process of water bloom prediction.

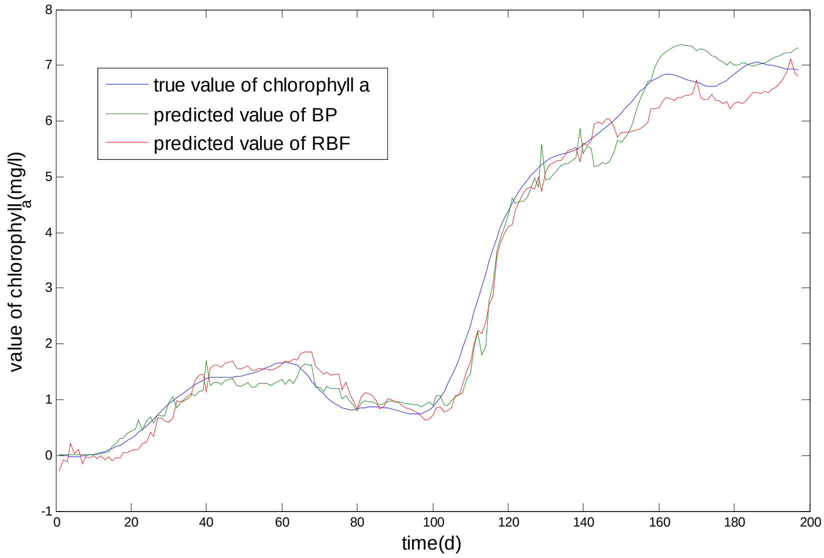

Fig. 3. Neural network prediction results.

The average relative error of short-term dynamic prediction of algae bloom based on RBF neural network is 0.41 %, while the value of BP neural network is 6.64 %. Simulation results show that this method could effectively realize short-term prediction of algae bloom, and it has higher prediction precision compared with BP neural network.

4. Summary

This paper proposes a method of short-term dynamic prediction of algae bloom based on artificial neural network, which uses principal component analysis to extract the principal components that can characterize characteristics of the formation of algae bloom. The simulation results show that the predicted results were more accurate, and the method is able to judge the trend of the blooms outbreak.

Acknowledgements

This study is supported by The National Natural Science Foundation of China (51179002), The Innovation Ability Enhance Plan of Beijing Municipal University (PXM2013-014213-000098), The Science and Technology Innovation Platform of Beijing Municipal Commission of Education (PXM2013_014213_000044) and Graduate Students Scientific Research and Academic Innovation Fund Project of Beijing Technology and Business.

References

[1]. Ding Ling, Pang Yong, Li Ling, Gao Guang, Simulation study on algal dynamics under different hydrodynamic conditions, *Acta Ecologica Sinica*, Vol. 25, Issue 8, 2005, pp. 1863-1868.

[2]. Ren Hongxiang, Zhang Daijun, Lu Peili, Xu Danyu, The eutrophication model of the lakes and reservoirs based on the structured programming, *Environmental Pollution & Control*, Issue 5, 2005, pp. 371-374+381-318.

[3]. Pei Hongping, Luo Nina, Jiang Yong, Application of back propagation neural network for predicting the concentration of chlorophyll_a in West Lake, *Acta Ecologica Sinica*, Vol. 24, Issue 2, 2004, pp. 246-251.

[4]. Li Dagang, Wang Xiaoyi, Liu Zaiwen, Xu Jiping, Zhao Xing, Dai Jun, Research on water-bloom prediction based on process neural network, *Computers and Applied Chemistry*, Issue 2, 2011, pp. 173-176.

[5]. Liu Zaiwen, Lv Siying, Wang Xiaoyi, Cui Lifeng, Huang Zhenfang, Forecast methods for algal bloom in rivers and lakes, *Water Resources Protection*, Issue 5, 2008, pp. 42-47.

[6]. Gao Jixi, Duan Feizhou, Xiang Bao, The application of principal component analysis to agriculture soil contamination assessment, *Geographical Research*, Issue 5, 2006, pp. 836-842.

[7]. Cao Yingui, Zhou Wei, Wang Jing, Yuan Chun, Zhao Li, Comparative on regional cultivated land intensive use based on principal component analysis and analytic hierarchy process in Three Gorges reservoir area, *Transactions of the Chinese Society of Agricultural Engineering*, Issue 4, 2010, pp. 291-296.

[8]. Zhang Yumei, Qu Shiru, Wen Kaige, A short-term traffic flow forecasting method based on chaos and RBF neural network, *Systems Engineering*, Issue 11, 2007, pp. 26-30.

[9]. Ye Jian, Ge Lin-Dong, Wu Yue-Xian, An application of improved RBF neural network in modulation recognition, *Acta Automatica Sinica*, Vol. 33, Issue 6, 2007, pp. 652-654.

[10]. Liu Yinhu, Li Shaoming, Single neuron PID control based on dynamic RBF neural network on-line identification, *Journal of System Simulation*, Issue S2, 2006, pp. 804-807.
