Computers and Electronics in Agriculture 150 (2018) 257–265

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework

Juan Huan[a], Weijian Cao[a], Yilin Qin[a,b,*]

[a] School of Information Science & Engineering, Changzhou University, Changzhou 213164, China
[b] Changzhou Technical Institute of Tourism & Commerce, Changzhou 213023, China

ARTICLE INFO

Keywords:
Predicted dissolved oxygen
Ensemble empirical mode decomposition
Least squares support vector machine
Bayesian evidence framework
Single point iterative method

ABSTRACT

In order to improve the accuracy and effectiveness of dissolved oxygen (DO) prediction, a combined forecasting model based on ensemble empirical mode decomposition (EEMD) and least squares support vector machine (LSSVM) is proposed. Firstly, the DO time series are decomposed into a group of relatively stable subsequences by ensemble empirical mode decomposition to reduce mutual influences among diverse trend information. Secondly, the decomposed subsequence is reconstructed by phase space reconstruction (PSR), and then, an LSSVM optimized by the Bayesian evidence framework prediction model of each sub-sequence is established. Lastly, we use Bp neural network to reconstruct the predicted values of each component to obtain the predicted value of the original DO sequence. This paper used the single point iterative method to achieve multi-step prediction in order to obtain forecasting results for 24 h into the future. EEMD-LSSVM is tested and compared with other algorithms in the Jiangsu Liyang huangjiadang special aquaculture farms. The experimental results show that the proposed combination prediction model of EEMD-LSSVM has a better prediction effect than WD-LSSVM, EEMD-ELM and standard LSSVM methods. The relative mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE) and the largest error (e_{max}) for the EEMD-LSSVM model are 0.0261, 0.2161, 0.1721 and 0.0767, respectively. Consequently, it is clear that the EEMD-LSSVM model has high forecast accuracy and generalization ability.

1. Introduction

Information technology has become an important tool for the sustainable development of modern aquaculture. In all aspects of aquaculture, such as intelligent feeding, disease monitoring and diagnosis, water quality monitoring, forecasting and early warning, information technology is widely used. Dissolved oxygen (DO) is one of the key water quality parameters for water products, which reflects changes in water quality in aquaculture. The water quality of an aquaculture pond has a direct impact on the growth of aquatic animals and product quality (Hu et al., 2015). For more than 16 h in a continuous 24-h period, the dissolved oxygen value must be greater than 5 mg/L, and for the remaining time, it should not be less than 3 mg/L. When dissolved oxygen is less than 3 mg/L, it will have a great impact on the feeding, digestion and health of fish. It is necessary to make a 24 h, or even longer, prediction of dissolved oxygen, which is convenient for managers in making early decisions. Therefore, accurate prediction of dissolved oxygen has very important economic value and practical significance.

In recent years, many studies have focused on dissolved oxygen (DO) content prediction and have obtained significant results. Liu et al. (2013) studied the prediction of DO content based on least squares support vector regression optimized by improved particle swarm optimization and later proposed a hybrid dissolved oxygen content forecasting model based on wavelet analysis (WA) and CPSO-LSSVR (Liu et al., 2014). Huan and Liu (2016) designed an approach using K-means clustering and ELM neural network to predict DO. The artificial neural network (ANN) and hybrid wavelet-ANN (WANN) models were used to predict thirty-minute dissolved oxygen levels in the River Calder (Ravansalar et al., 2015). Ahmed (2014, 2015) used artificial neural networks to forecast the DO and applied the adaptive neurofuzzy inference system to estimate the DO of the Surma River. Khan and Valeo, 2017 used a possibility theory-based fuzzy neural network to predict DO. Chen et al. (2016) presented a three-dimensional prediction model for DO content based on the PSO-BPANN algorithm coupled with Kriging Interpolation. The study of Arya and Zhang (2015) proposed an ARMA prediction model based on wavelet decomposition, which was used to predict changes in the DO and temperature in the Stillaguamish

* Corresponding author at: China Changzhou University, Gehu Road No. 1, Wujin, Changzhou, Jiangsu 213164, China.
E-mail address: qinyilin101@163.com (Y. Qin).

https://doi.org/10.1016/j.compag.2018.04.022
Received 4 January 2018; Received in revised form 21 March 2018; Accepted 21 April 2018

River.

However, several previous studies using a single model for DO forecasting usually directly applied the original DO data to construct forecasting models. Because of the nonlinearity and nonstationarity of DO data, it is difficult to describe the tendency of DO and to improve prediction accuracy. In order to establish a suitable and effective prediction model, the original data characteristics of DO need to be considered and analyzed. Therefore, empirical mode decomposition (EMD), as a special adaptive and direct data processing method developed for dealing with nonlinear and nonstationary data, is used to decompose the original DO data into a sum of intrinsic mode function (IMF) components and one residue, which can improve the accuracy of forecasting. The EMD technique, however, also suffers from a number of drawbacks such as mode mixing and end effects. Zhaohua and Huang (2009) proposed an improved decomposition technique, which is an ensemble empirical mode decomposition (EEMD) technique, to alleviate the mode-mixing problem. This method can effectively reflect the essence of the original signal and has been quickly adopted in many fields. For example, Wang et al. (2014) used a combined independent component analysis (ICA) and EEMD to separate the fault signal from a mixing signal. Xu et al. (2016) proposed ensemble empirical mode decomposition and an improved artificial bee colony algorithm model for prediction of pH values. Yu et al. (2015) presented a cayenne pepper height forecasting method based on ensemble empirical mode decomposition (EEMD) and the Elman neural network (ELM).

The rest of this paper is organized as follows: Section 2 describes EEMD, LSSVM, and Bayesian evidence framework methodologies. Section 3 presents the detailed modeling steps of the EEMD-LSSVM in a Bayesian evidence framework model. Section 4 gives the actual data and compares the performance of the proposed model with other prediction models. Section 5 discusses the conclusions of this article.

2. Materials and methodology

2.1. Study area data source

This paper selected the Liyang huangjiadang special aquaculture farms in Changzhou city, Jiangsu province, China, as a test area. The experiment pond was approximately 1.53 km^2 in size, and the water level averaged 1.5–2.5 m. The pond has a pond circulating water system, equipped with a dissolved oxygen sensor, oxygen pump, wireless monitoring system and other modern fishery equipment. Aquaculture environment data were obtained from the aquaculture remote wireless monitoring system, and the system collected data every hour online. It is convenient for us to obtain DO data from the remote Internet system management platform. The system topology structure diagram is shown in Fig. 1

2.2. EMD and EEMD

Empirical mode decomposition (EMD), first proposed by Huang et al. (1998), is a novel empirical analysis tool used for processing nonlinear and nonstationary datasets. The main function of EMD is to decompose DO time series into a set of several simple intrinsic mode function (IMF) components and one residue.

Let $x(t)$ be a given original time series. The detailed steps of EMD calculation can then be described as follows (Zhao et al., 2017).

Step 1. Determine all the maximum and minimum values for the entire time series $x(t)$.

Step 2. Calculate the upper envelope, which can be derived by connecting all the local maxima using a cubic spline line. The lower envelope can also be obtained with this method. Calculate the mean of the upper and lower envelopes, referred to as $n_1(t)$. Calculate the first difference $z_1(t)$ between the original series data $x_1(t)$ and $n_1(t)$:

$$z_1(t) = x_1(t) - n_1(t) \tag{1}$$

Step 3. Check whether $z_1(t)$ satisfies the requirements of IMF. If $z_1(t)$ does not satisfy an IMF, replace $x(t)$ with $z_1(t)$ and repeat Step 2 until the termination criterion is satisfied. The termination criterion is as follows:

$$\sum_{t}^{l} \frac{[z_{i-1}(t) - z_i(t)]^2}{[z_{i-1}(t)]^2} \leqslant \eta \tag{2}$$

where l is the length of the $z_i(t)$, η is the terminated parameter, and i is the number of iterations. η is assumed to be in the range of 0.2–0.3 (Yu et al., 2015). This paper uses $\eta = 0.2$.

After repeating step 2 k times, we obtain the difference value $z_k(t)$ that satisfies the requirements of IMF. Let $q_1(t) = z_k(t)$.

However, if $z_1(t)$ is an IMF, then $z_1(t)$ is denoted as the first IMF $q_1(t)$ and $x(t)$ is replaced with the residue $c_1(t)$:

$$c_1(t) = x(t) - q_1(t) \tag{3}$$

Step 4. $c_1(t)$ is replaced by x(t), and Steps 1–3 are repeated. We can then obtain the rest

of the IMF and a trend of $r(t)$. After the EMD calculation, the original time

series data $x(t)$ can be decomposed into the sum of all the IMF components and a residue as follows:

$$x(t) = \sum_{i=1}^{m} q_i(t) + r(t) \tag{4}$$

EEMD effectively uses noise characteristics to reduce mode aliasing. The EEMD signal decomposition steps are as follows:

Step 1. Add a white noise sequence, which should be subject to a

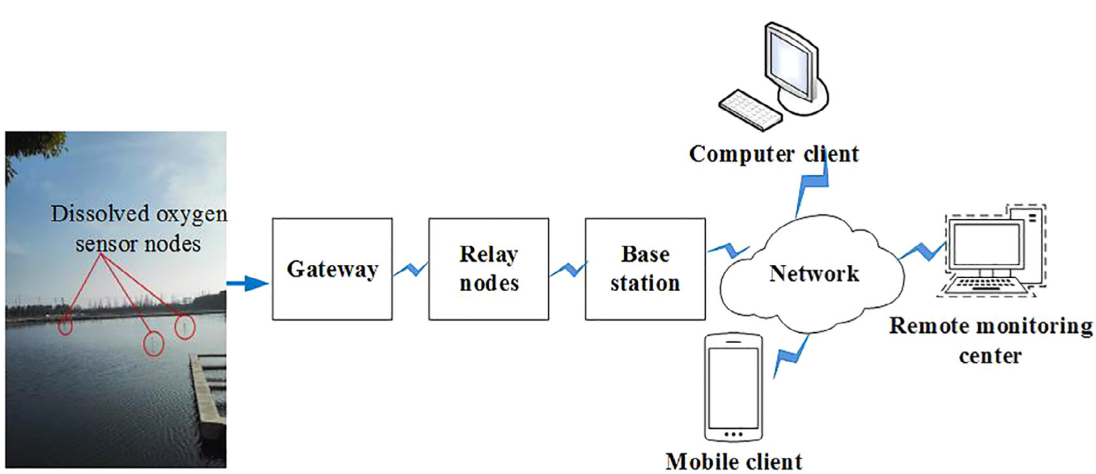

Fig. 1. The system topology structure diagram.