Engineering Applications of Artificial Intelligence 29 (2014) 114–124

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture

Shuangyin Liu [a,b,c,d], Longqin Xu [a], Yu Jiang [b,c,d], Daoliang Li [b,c,d,*], Yingyi Chen [b,c,d], Zhenbo Li [b,c,d]

[a] College of Information, Guangdong Ocean University, Zhanjiang, Guangdong 524025, China
[b] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[c] Beijing ERC for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China
[d] Beijing ERC for Advanced Sensor Technology in Agriculture, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 23 June 2013
Received in revised form
26 September 2013
Accepted 27 September 2013
Available online 28 November 2013

Keywords:
Least squares support vector regression
Wavelet analysis
Cauchy particle swarm optimization algorithm
Dissolved oxygen content forecasting
Parameter optimization

ABSTRACT

To increase prediction accuracy, reduce aquaculture risks and optimize water quality management in intensive aquaculture ponds, this paper proposes a hybrid dissolved oxygen content forecasting model based on wavelet analysis (WA) and least squares support vector regression (LSSVR) with an optimal improved Cauchy particle swarm optimization (CPSO) algorithm. In the modeling process, the original dissolved oxygen sequences were de-noised and decomposed into several resolution frequency signal subsets using the wavelet analysis method. Independent prediction models were developed using decomposed signals with wavelet analysis and least squares support vector regression. The independent prediction values were reconstructed to obtain the ultimate prediction results. In addition, because the kernel parameter δ and the regularization parameter γ in the LSSVR training procedure significantly influence forecasting accuracy, the Cauchy particle swarm optimization (CPSO) algorithm was used to select optimum parameter combinations for LSSVR. The proposed hybrid model was applied to predict dissolved oxygen in river crab culture ponds. Compared with traditional models, the test results of the hybrid WA–CPSO-LSSVR model demonstrate that de-noising and capturing non-stationary characteristics of dissolved oxygen signals after WA comprise a very powerful and reliable method for predicting dissolved oxygen content in intensive aquaculture accurately and quickly.

1. Introduction

Predicting the dissolved oxygen content in modern intensive aquaculture can provide a basis for water quality control and management decisions that can reduce aquaculture risks and optimize operation (Han et al., 2011). Dissolved oxygen content control and management in intensive aquaculture is crucial for growing organisms and has a significant impact on the quantity and quality of the final product (Carbajal-Hernández et al., 2012). However, dissolved oxygen content forecasting is a dynamic, complex and nonlinear system. Modeling such a complex system involves many uncertainties originating from the randomness of meteorology, hydrodynamics, water chemistry, human activity and biological factors (Liu et al., 2012a,b). Some numerical or physical model studies of water quality forecasting have been conducted; however, they were not successful enough in forecasting the dissolved oxygen content due to incomplete background information, incomplete monitoring data, inaccurate initial conditions, and limited spatial resolution (Kisi and Cimen, 2011). Recently, artificial intelligence (AI) techniques are beginning to be able to imitate this behavior and overcome the drawbacks.

Artificial intelligence techniques are an efficient alternative tool for complex and nonlinear system modeling in many disciplines (Hatzikos et al., 2005; Preis and Ostfeld, 2008; Mahapatra et al., 2011; Faruk, 2010; Tan et al., 2012; Liu et al., 2012a,b, 2013a,b). Although the models usually do not consider internal mechanisms, they can predict water quality via the relationships between inputs and outputs and have been successfully used to predict water quality. Hatzikos et al. (2005) utilized neural networks with active neurons as a modeling tool for the prediction of seawater quality indicators, such as water temperature, pH, dissolved oxygen and turbidity, but is has the drawback of having a large number of control parameters, making it difficult to find a stable solution. Faruk (2010) proposed a hybrid artificial neural networks (ANN) and autoregressive integrated moving average (ARIMA) model to predict a time series of water quality data. However, determining the number of hidden layers and hidden neurons per layer in this hybrid model is difficult. Also fitting the model to data suffers from local minima and over fitting. Han et al. (2011) presented a flexible

* Corresponding author at: China Agricultural University, P.O. Box 121, 17 Tsinghua East Road, Beijing 100083, China. Tel.: +86 10 62736764; fax: +86 10 62737741.
E-mail addresses: dliangl@cau.edu.cn, hdlsyxlq@126.com (D. Li).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.09.019

structure radial basis function (RBF) neural network for water quality prediction. The hidden neurons in the RBF neural network can be added or removed online based on neuron activity and mutual information (MI). Experimental results demonstrated its effectiveness. Liu et al. (2012a,b) analyzed the important factors for predicting dissolved oxygen in *Hyriopsis cumingii* ponds and used Elman neural networks to forecast the dissolved oxygen content in aquaculture of *H. cumingii*. Their approach, which uses a steepest descent search, converges slowly and does not support non-smooth models.

Least squares support vector regression (LSSVR) simplified the model standard SVR to a great extent by applying linear least squares criteria to the loss function instead of a traditional quadratic programming method, which greatly improves the calculation speed and accuracy (Suykens et al., 2002; Vapnik, 2000; Liao et al., 2011; Rumpf et al., 2012). Compared with artificial neural networks, LSSVR seeks to minimize the upper bound of the generalization error instead of the empirical error and can provide more reliable and better generalization performance under the same training conditions (Liao et al., 2011; Singh et al., 2011). LSSVR has been successfully applied to many fields of function approximation and pattern recognition. However, LSSVR performance heavily depends on the choice of the kernel parameter σ and the regularization parameter γ, which are necessary to define the optimization problem and the final LSSVR model. Unfortunately, there is no exact method to obtain the optimal set of parameters (σ, γ) in the LSSVR model. Therefore, some researchers applied intelligent search algorithms to optimize these parameters. They include analytical techniques, heuristic techniques, simulated annealing algorithms, genetic algorithms, particle swarm optimization (PSO) algorithms, and additional evolutionary strategies (Zhou et al., 2012; Pahasa and Ngamroo, 2011; Vong et al., 2006; Lin et al., 2008; Liu et al., 2012a,b; Babaoglu et al.,2010; Abdi and Giveki, 2013), which can be applied to implement a robust search of the parameter (σ, γ) search space. For example, Liu et al. (2013a,b) presented the dissolved oxygen prediction model based on least squares support vector regression optimized by improved particle swarm optimization, and acquired a satisfactory forecasting result. Compared with other heuristic algorithms, the PSO algorithm is considered an excellent technique to solve the combinatorial optimization problem. PSO possesses the capability to escape from local optima, is easily implemented, and has fewer parameters to be set (Clere and Kennedy, 2002; Liang et al., 2006; Hsieh et al., 2012; Wu et al., 2013). In this study, the improved particle swarm optimization algorithm was applied to optimize the (σ, γ) parameters of the LSSVR.

The aforementioned study efforts were usually based on the assumption that the original data should be reliable and accurate (Najah, 2012). However, data usually contain systematic and random errors, even when acquired by advanced instruments (Wang and Shi, 2013). For example, there are many sources of noise in water quality data collection systems, such as sensor calibration and aging caused by water erosion and electric power fluctuations. Unfortunately, this noise corrupts the data and distorts the results of forecasting models. Therefore, it is necessary to remove such noise from the original data.

Wavelet analysis is called a 'microscope' in mathematics and can distinguish between noise and useful signals. It is able to capture the non-stationary characteristics of dynamic systems and has been successfully applied to knowledge discovery and pattern recognition (Wu and Law, 2011; Eynard et al., 2011; Wang and Shi, 2013; Najah et al., 2012; Liang et al., 2012; Kisi and Cimen, 2011; Kao et al., 2013; Kalteh, 2013). Najah et al. (2012) proposed an augmented wavelet de-noising technique with a Neuro-Fuzzy Inference System (WDT-ANFIS) based on a data fusion module for water quality prediction. Wu and Law (2011) proposed a robust wavelet support vector machine (RW ν-SVM) based on wavelet theory and a modified support vector machine for car demand forecasting. The simulation results demonstrated the effectiveness of this algorithm. Liang et al. (2012) explored a hybrid method that coupled a discrete wavelet transform with a support vector machine based on amino acid polarity to predict the subcellular localizations of prokaryotic and eukaryotic proteins. Kao et al. (2013) proposed a stock price-forecasting model that integrates a wavelet transform, a multivariate adaptive regression spline, and support vector regression. This model can address the problem of wavelet sub-series and improves accuracy simultaneously. Kalteh (2013) investigated the relative accuracy of artificial neural network and support vector regression models coupled with wavelet transform for monthly river flow forecasting and found that SVR models coupled with wavelet transforms provide better forecasting performance than ANN models coupled with wavelet transform. These studies all showed that wavelet analysis is an effective tool for precisely locating irregularly distributed multi-scale features of water quality indexes in space and times. Though there are many different types of AI techniques in artificial neural networks, support vector machines and least squares support vector regressions (LSSVR) have recently gained attention in the literature. Nevertheless, few of these studies applied these methods to water quality management or water quality forecasting in modern intensive aquaculture. In this study, we used wavelet analysis to de-noise and extract features of the original dissolved oxygen data to improve forecast accuracy.

LSSVR is a very successful and efficient technique for overcoming prediction problems with small samples, nonlinear samples, high-dimensional data and data with local minima. Wavelet analysis is able to de-noise and capture the non-stationary characteristics of the dynamic systems. Based on the above factors, this paper proposes a new hybrid prediction model for dissolved oxygen content in modern intensive aquaculture combining wavelet analysis, least squares support vector regression and Cauchy particle swarm optimization (CPSO) algorithms. The CPSO method was used to optimize the kernel parameter σ and regularization parameter γ of LSSVR. Thus, the hybrid model can improve the overall performance of prediction in both linear and nonlinear patterns. WA–CPSO-LSSVR was tested and compared with other algorithms using the dissolved oxygen data in river crab ponds. The results show that the accuracy of prediction and the capability of generalization are greatly improved by our proposed approach.

The rest of this paper is organized as follows: Section 2 reports the construction of a hybrid forecasting model based on the wavelet analysis approach, least squares support vector regression and Cauchy particle swarm optimization. In section 3, the steps of developing a hybrid prediction method are described. Section 4 describes an application of the hybrid forecasting model based on WA–CPSO-LSSVR. Section 5 draws conclusions and future works.

2. Materials and methods

2.1. Study area data source

The data used in this study were produced by the Digital Wireless Monitoring System of Aquaculture Water Quality. The system was installed at the zone of technology application and demonstration at China Agricultural University, at the Yixing base of intelligent aquaculture management systems in Jiangsu province, China. The system structure diagram is shown in Fig. 1. The basic structure of the system can be divided into three major parts: the data apperception layer; the transport layer; and the application layer for data acquisition, reliable transmission, intelligent information processing, and logical operation (Liu et al., 2012a,b).