Mathematical and Computer Modelling 58 (2013) 458–465

Contents lists available at SciVerse ScienceDirect

Mathematical and Computer Modelling

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

A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction

Shuangyin Liu[a,b], Haijiang Tai[a], Qisheng Ding[a], Daoliang Li[a,*], Longqin Xu[b], Yaoguang Wei[a]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] College of Information; Guangdong Ocean University, Zhanjiang Guangdong 524025, China

ARTICLE INFO

Article history:
Received 25 October 2011
Accepted 8 November 2011

Keywords:
Water quality prediction
Support vector regression
Genetic algorithms

ABSTRACT

Water quality prediction plays an important role in modern intensive river crab aquaculture management. Due to the nonlinearity and non-stationarity of water quality indicator series, the accuracy of the commonly used conventional methods, including regression analyses and neural networks, has been limited. A prediction model based on support vector regression (SVR) is proposed in this paper to solve the aquaculture water quality prediction problem. To build an effective SVR model, the SVR parameters must be set carefully. This study presents a hybrid approach, known as real-value genetic algorithm support vector regression (RGA–SVR), which searches for the optimal SVR parameters using real-value genetic algorithms, and then adopts the optimal parameters to construct the SVR models. The approach is applied to predict the aquaculture water quality data collected from the aquatic factories of YiXing, in China. The experimental results demonstrate that RGA–SVR outperforms the traditional SVR and back-propagation (BP) neural network models based on the root mean square error (RMSE) and mean absolute percentage error (MAPE). This RGA–SVR model is proven to be an effective approach to predict aquaculture water quality.

1. Introduction

Aquaculture water is an important aspect of the river crab's habitat in the intensive breeding of river crab, and the water quality determines the growth status and product quality directly. Once the water quality deteriorates and the crabs are in a poor environment, it is very easy for there to be an outbreak of some diseases; also there is the decline in the quality and even a large number of dead river crabs in a short time, which will cause great economic losses to the farmers if remedial measures are not taken in a timely manner. So, taking advantage of modern information technology to have early warnings of water conditions and enable the dynamic change of water is an urgent and important matter.

Aquaculture water is an open, nonlinear, dynamic, complex system. Water quality is affected by many factors such as physics, chemistry, hydraulics, biology, meteorology, and human activities, and the water quality parameters are nonlinear, time varying, random and delayed, because of the interactions between them. Thus, it is difficult to describe them quantitatively using accurate mathematical models and to establish an accurate, perfect, nonlinear prediction model using traditional methods.

Prediction of water quality focuses mainly on lakes, rivers, reservoirs, estuaries, and other large expanses of water using the gray system theory, neural networks, statistical analysis methods, time series models, both in China and elsewhere.

* Corresponding author. Tel.: +86 10 62736764; fax: +86 10 62737741.
E-mail address: dliangl@cau.edu.cn (D. Li).

0895-7177/$ – see front matter
doi:10.1016/j.mcm.2011.11.021

Partalas et al. studied the greedy ensemble selection family of algorithms for ensembles of regression models to solve the forecasting of water quality [1]; Feifei Li et al. established back-propagation (BP) and autoregressive (AR) versions of the short-term forecasting model to predict dissolved oxygen [2], Eun Hye Naa et al. designed a dynamic three-dimensional water quality model to predict phytoplankton growth patterns in time and space [3]; Han has presented a flexible structure radial basis function neural network (FS-RBFNN) to predict the wastewater biochemical oxygen demand (BOD) index [4].

Palani et al. developed a neural network model to forecast the amount of dissolved oxygen in seawater [5]; Bikash Sarkar proposed a water quality model to predict the changes of temperature in an indoor fish pond [6]; Yu Deng adopted a wavelet neural network model based on wavelet theory and neural network theory to forecast the drinking water permanganate index [7]. However, neural networks suffer from a few weaknesses, which include the need for numerous controlling parameters, difficulty in obtaining a stable solution, and the danger of over-fitting.

Support vector regression (SVR) is a novel learning machine based on statistical learning theory and a structural risk minimization principle, which has been successfully used for nonlinear system modeling [8]. Yunrong Xiang employed a least squares support vector machine (LS-SVM) and a particle swarm optimization model to predict the quality of a drinking water source [9]. Compared with artificial neural networks, an SVM provides more reliable and better performance under the same training conditions [10,11]. Although it has excellent features, SVR is limited in academic research and industrial applications, since the user must define various parameters appropriately. The SVR parameters must be set carefully in order to construct the SVR model efficiently [12–14]. Inappropriately chosen SVR parameters will result in over-fitting or under-fitting, and different parameter settings may also cause significant differences in performance [15]. Thus, selecting the optimal parameters is an important step in SVR design. However, no general guidelines are available to help in selecting these parameters [16–18]. So, we propose a hybrid approach of SVR with real-value genetic algorithm (RGA) optimization is developed by adopting an RGA to determine the SVR free parameters, and so the generalization ability and forecasting accuracy are improved in this study. The approach is used to forecast water quality in a high-density crab culture situation. The traditional SVR model and a BP neural network were also investigated for comparison. The experimental results show that an improvement in predictive accuracy and capability of generalization can be achieved by our proposed approach.

The structure of the paper is as follow. In Section 2, we introduce the real-value genetic algorithm (RGA) and support vector regression (SVR), and then the hybrid RGA–SVR model is proposed. Section 3 describes the data source and experimental setting and explains the process for determining the parameters of the RGA and SVR models. Section 4 discusses the results and analysis of the hybrid RGA–SVR model used in on-site aquaculture water quality prediction. Section 5 concludes the study, and suggests directions for future investigations.

2. Methodology

2.1. Genetic algorithms (GAs)

GAs are stochastic search techniques that can search large and complicated spaces using ideas from natural genetics and the evolutionary principle [19,20]. The idea of this method, which was inspired by the theory of natural evolution, was first proposed by Holland [21]. A genetic algorithm works with a population of individual strings (chromosomes), each representing a possible solution to a given problem. Each chromosome is assigned a fitness value according to the result of the fitness function. Highly fit chromosomes are given more opportunities to reproduce and the offspring share features taken from their parents. The GA is a simple but powerful tool for finding the global solution to an optimization problem. It is suitable for large-scale and complex nonlinear optimization problems, and it has the tendency to find the global optimal solution [22].

The procedure of a GA can be summarized in the following steps.

1. Choose a randomly generated population.
2. Calculate the fitness of each chromosome in the population.
3. Create the offspring by genetic operators: selection, crossover, and mutation.
4. Judge the stopping criteria. If the stopping criteria are met, the genetic algorithm would be stopped. Otherwise, repeat steps 2–4 using the generated offspring as the new starting population.

There are two types of coding method for GAs. They are real and binary-coded GAs. However, binary coding has some disadvantages, such as needing more memory, not being so flexible, and requiring a large amount of decode computation, so it is difficult to solve large-scale multi-parameter optimization problems using a binary coding genetic algorithm in a small-memory computer [23].

In contrast to the binary genetic algorithm (BGA), the real-value genetic algorithm (RGA) uses a real value as a parameter of the chromosomes in the population without the coding and encoding process prior to calculating the fitness value [24]. Thus, the RGA is more straightforward, faster, and more efficient. GAs have been used in a number of applications in engineering and social science. They have recently been applied, for example, to the optimization of the parameters of support vector machines for predicting bankruptcy [23], parallel searching for an optimal feature subset [25], and efficient selection and assignment of material handling equipment [26].