Computers and Electronics in Agriculture 157 (2019) 329–338

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Prediction of dissolved oxygen content in aquaculture using Clustering-based Softplus Extreme Learning Machine

Pei Shi[a,b], Guanghui Li[a,c,*], Yongming Yuan[b], Guangyan Huang[d], Liang Kuang[e]

[a] School of IoT Engineering, Jiangnan University, Wuxi 214122, China
[b] Freshwater Fisheries Research Center of Chinese Academy of Fishery Sciences, Wuxi 214081, China
[c] Engineering Research Center of Internet of Things Technology Applications, Ministry of Education, Wuxi 214122, China
[d] School of Information Technology, Deakin University, Melbourne, Australia
[e] School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China

ARTICLE INFO

Keywords:
Time series
Dissolved oxygen prediction
Clustering
Extreme learning machine

ABSTRACT

Accurate and efficient prediction of dissolved oxygen from time series data is critical for aquaculture that needs intelligent management and control. However, data streams of dissolved oxygen that are nonlinear and continuously generated challenge the existing prediction methods. This paper provides a novel Clustering-based Softplus Extreme Learning Machine method (CSELM) to accurately and efficiently predict dissolved oxygen change from time series data. The CSELM adopts *k*-medoids clustering to group the dataset into different clusters based on Dynamic Time Warping (DTW) distance, and uses a new Softplus ELM algorithm to discover a common trend in a cluster of time series pieces (within the same period) and then predict the future trend. The Softplus ELM improves ELM using a new activation function, Softplus, to solve the nonlinear and continuous problems of time series data streams and adopting partial least squares (PLS) to avoid the instability of output weight coefficients. The DTW based clustering in CSELM improves the efficiency while tolerating some data loss and uncertain outliers of sensor time series. Softplus based on PLS optimizes the performance and increases the accuracy of ELM. We have demonstrated that CSELM achieves better prediction results than PLS-ELM and ELM models in terms of accuracy and efficiency in a real-world dissolved oxygen content prediction.

1. Introduction

With the advance of the internet of things (IoT) monitoring systems, water quality time series data that are continuously generated and daily accumulated into a large volume provide abundant information to assist prediction. For example, we can learn the temporal trends from the water quality time series data during different periods of a day and use these common trends to predict the water quality changes in the near future. Dissolved oxygen is an important water quality index in intensive aquaculture that determines the growth status of tilapia (Zhu et al., 2010). Since sensor fault occurs often in reality (Yan et al., 2018), accurate and efficient prediction of dissolved oxygen from time series data is critical for water quality management and control, which can minimize the aquaculture risks and increase the efficiency of fish culture.

Prediction of water quality change is a complex task, involving multiple parameters and dynamic delay process (Antanasijević et al., 2013; Bartzanas et al., 2013) and it is hard to complete efficiently. Also, the quality of the sensor data that is often reduced by data loss and uncertain outlier challenges the prediction accuracy (Liu et al., 2013). We need to develop prediction algorithms that can tolerate some degree of sensor data quality problems.

Unfortunately, existing water quality prediction methods cannot resolve the above two problems well. Among them, Artificial Intelligence (AI) is the most widely used technique for water quality prediction (Hatzikos et al., 2005; Mahapatra et al., 2011; Faruk, 2010). The AI-based techniques can be used like a black box, in which the relationships of data can be obtained without understanding their internal mechanisms, and they can achieve satisfied speed when applied to water quality prediction. Support Vector Machine (SVM) and Artificial Neural Network (ANN) are two typical methods for water quality prediction. Least squares support vector machine (LS-SVM) prediction model was used by Tan et al. (2012), and all of its parameters were determined empirically or randomly. A hybrid particle swarm algorithm optimized least squares support vector regression model was presented by Liu et al. (2013) to predict dissolved oxygen content. Yu

* Corresponding author at: School of Internet of Things (IoT) Engineering, Jiangnan University, Wuxi, China.
E-mail addresses: ship@ffrc.cn (P. Shi), ghli@jiangnan.edu.cn (G. Li).

https://doi.org/10.1016/j.compag.2019.01.004
Received 15 July 2018; Received in revised form 10 December 2018; Accepted 7 January 2019
Available online 11 January 2019
0168-1699/

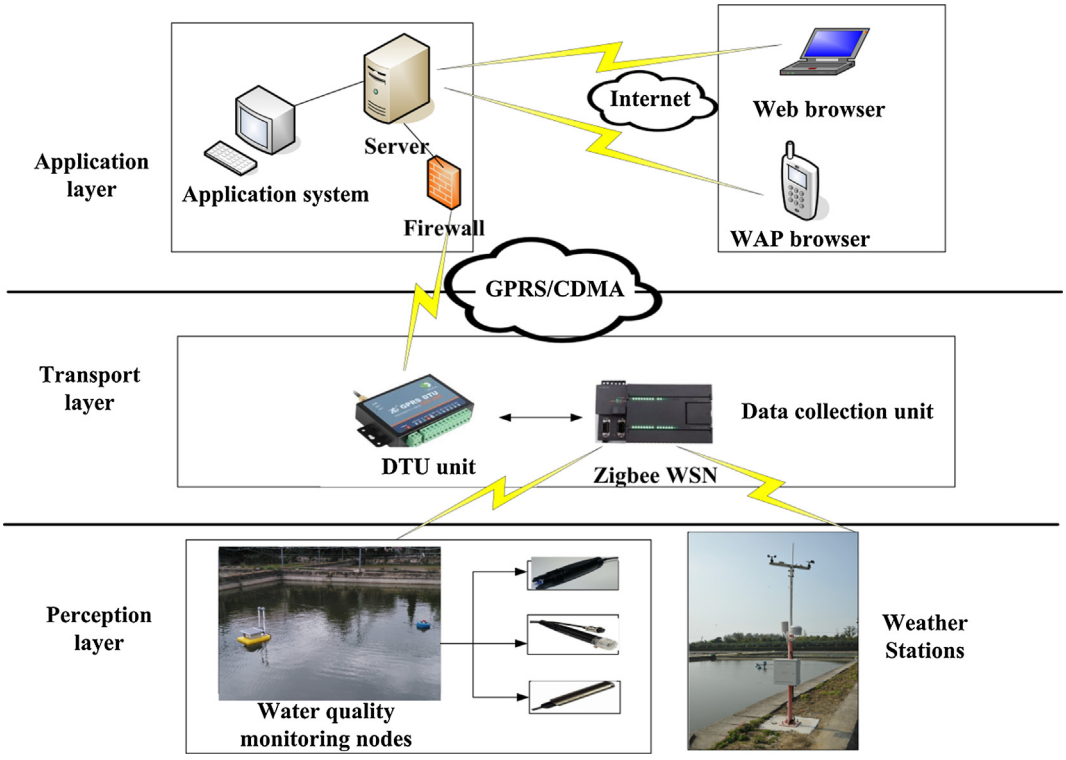


Fig. 1. The IoT monitoring system for collecting aquaculture water quality time series.

et al. (2016) proposed a hybrid RBFNN-IPSO-LSSVM model for water quality prediction, which utilized the improved Particle Swarm Optimization (IPSO) to determine the optimal parameters for Least Squares Support Vector Machine (LSSVM). However, SVM and its optimized models (e.g., genetic algorithms) are sensitive to data loss, time-consuming for training (Chen and Yuan, 2014) and not robust in prediction performances due to the difficulty in choosing suitable kernel functions for various applications.

ANN is robust and tolerates some faults (Amid and Gundoshmian, 2017). However, it is often overfitted and time-consuming due to the complexity of network topology and data. More intelligent algorithms have been developed to overcome ANN's limitations (Shahin, 2016). Palani et al. (2008) applied the artificial neural network (ANN) models, such as General regression neural networks (GRNN), multilayer perceptron neural network (MLP) and Back Propagation model (BP) three hidden layers with different activation functions to predict water quality parameters using continuous weekly water quality variables. Faruk (2010) presented a hybrid approach, combining seasonal ARIMA model and neural network back propagation model to predict the water quality parameters monthly. The aforementioned ANN models have numerous drawbacks, which include poor stability, low generalization, over-fitting (Liu et al., 2013).

ELM was introduced by Huang et al. (2006) as a new neural network training model. Compared with traditional ANN models, ELM is more efficient because of its simple network structure (Zhou et al., 2016). In dealing with nonlinear dynamic system, ELM has been successfully applied to many fields of prediction and classification because of its high efficiency and generalization capabilities (Zhao et al., 2013; Junior and Backes, 2016; Heddam and Kisi, 2017; Yan et al., 2017). In spite of this, the ELM performance heavily depends on the choice of activation function. Moreover, the output weight coefficients are instable when the input nodes are redundant. To design an ELM prediction model in dissolved oxygen forecasting, one must choose an activation function and deal with the problem of solving output weight coefficients.

In aquaculture, water quality data streams and meteorological data have periodic pattern and trend changes. So many models of water quality evaluation and prediction do not just focus on the improvement of algorithms, but also explore the patterns of data streams. Wavelet transform method was utilized by Xu et al. (2013) to capture the feature of water quality data. Liu et al. (2014) segmented the time series data collected online to capture the feature of dissolved oxygen data. A novel statistic variable similarity of days with similar weather was defined by Huan et al. (2017), which is calculated by the Pearson correlation coefficient. Based on the similarity of days, monitoring data were classified into several parts with *k*-means method to find out the potential relation. However, there is no exact approach to explore the potential rule of water quality data streams and meteorological data in day and night.

In this study, a novel Clustering-based Softplus Extreme Learning Machine method (CSELM) is proposed to accurately and efficiently predict dissolved oxygen change from time series data. The CSELM adopts *k*-medoids clustering to group the dataset into different clusters based on Dynamic Time Warping (DTW) distance, and utilizes a new Softplus ELM algorithm to predict the future trend. Traditional ELM and PLS-ELM models are used as a comparison. The experimental results demonstrate that the accuracy and efficiency are greatly improved by CSELM model in a real-world dissolved oxygen prediction.

One advantage of CSELM is that it can tolerate sensor data quality problems using clustering method and achieve satisfied efficiency and accuracy. We define a novel variable to measure the similarity by separating the time series into two segments: day-stream and night-stream. The *k*-mediods clustering based on DTW distance can discover the periodic change patterns and trend changes in similar time slots and divide the dataset into different clusters; and intelligent grouping of data helps tolerate data loss and greatly improve the prediction accuracy and efficiency. Another advantage of CSELM is that the Softplus ELM improves the prediction accuracy, since it has more optimized network performance.

The paper is organized as follows: in Section 2, we present the study area, data acquisition and CSELM prediction model. In Section 3, we demonstrate the accuracy and efficiency of the forecasting results. In Section 4, we give concludes from this study.

2. Materials and methods

2.1. Data acquisition

We have used aquaculture water quality time series data collected by IoT monitoring system as shown in Fig. 1 for experiments. The IoT monitoring system includes three layers for data collection, data transport and data application. In the data collection layer, we use various types of sensors to collect water quality parameters (e.g., dissolved oxygen, pH and water temperature) and meteorological parameters (e.g., temperature, humidity, atmospheric pressure, carbon dioxide, illumination intensity, photosynthetically active radiation, radiance, wind speed and direction). In the application layer, we process the time series data transferred from collection layer via the transport layer.

All the data were obtained from the Nanquan breeding base located in Wuxi city, Jiangsu province. The total area of Nanquan tilapia breeding base was 9.5 acres. The depth of the pond was 1.5 m, and the dissolved oxygen sensor, pH sensor, water temperature sensor were all placed in 0.8 m underwater. These sensors were packed in a device under a float with a solar panel. The automatic meteorological station was installed on the shore of the pond. All these data were transferred to the IoT monitoring system for collecting aquaculture water quality and meteorological time series.

The water quality and meteorological time series data set for the experiments includes 4464 data sets (sampled 10 min once) from July 1 to July 31, 2016. The first 3780 data sets were used for training, and the remaining 684 sets were used for testing. Because the water quality in pond are mainly affected by physical and chemical factors, include dissolved oxygen, pH and water temperature, humidity, temperature, atmospheric pressure, carbon dioxide, illumination intensity, photosynthetically active radiation, radiance, wind speed and direction. These eleven influencing factors and the previous DO value form the input vectors of the forecasting model and the result of this model is the dissolved oxygen value of the next ten minute, and as shown in Fig. 2.

2.2. Overview of the proposed CSELM prediction model

The characteristics of dissolved oxygen time series data in day and night are quite different. Moreover, the dissolved oxygen curves have correlation with meteorological condition. For the purpose of discovering the periodic change patterns and trend changes between dissolved oxygen and the aquaculture environment, this paper used the *k*-medoids clustering based on DTW distance to divide the data samples into different clusters with meteorological index. Our CSELM model, as shown in Fig. 3, is a nonlinear prediction model for predicting the dissolved oxygen content. Before training the model, we have preprocessed the original data by improving the “bad data”. Then, *k*-medoids will be used to group data samples into several clusters by time series similarity measured by DTW. Finally, the prediction models in these clusters were built to realize the prediction of dissolved oxygen.

Sample preparation: Collect the real time water quality data and meteorological data by the IoT monitoring system. Input data of aquaculture environment factors to construct training set and testing set.

Parameter selection: Select activation function and determine the algorithm optimal parameters.

Model training and application: Train the model by training sets (3780 data samples) and obtain the CSELM prediction model. Test the CSELM model by testing sets (684 data samples). Form a model of input variables X_i and output estimation value Y. The detail of CSELM modeling is shown as Table 1.

In data preprocessing (in Step 1 of Algorithm 1), we normalize data and correct or remove anomalous or missing data. We used the factor analysis method (Ye et al., 2015) to calculate the meteorological index as reference for repairing anomalous data. Entropy method was used to

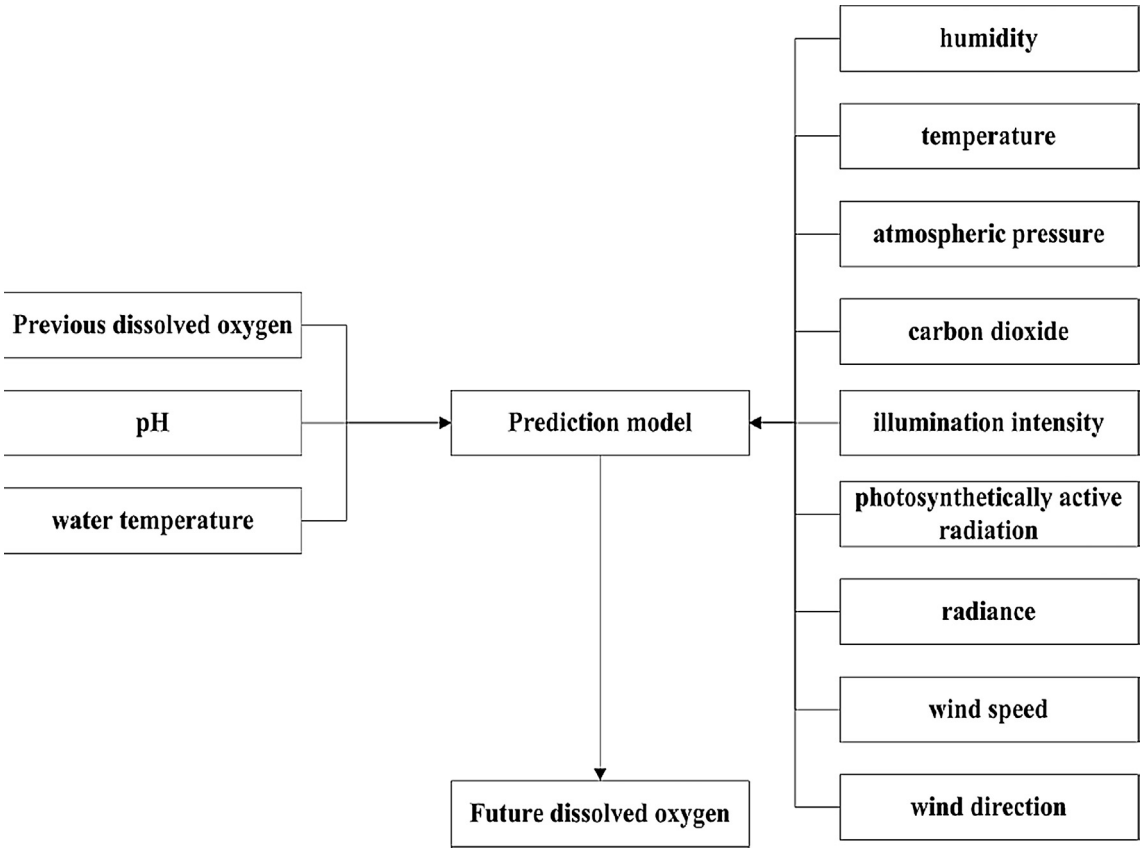


Fig. 2. Input-output structure of forecasting model.

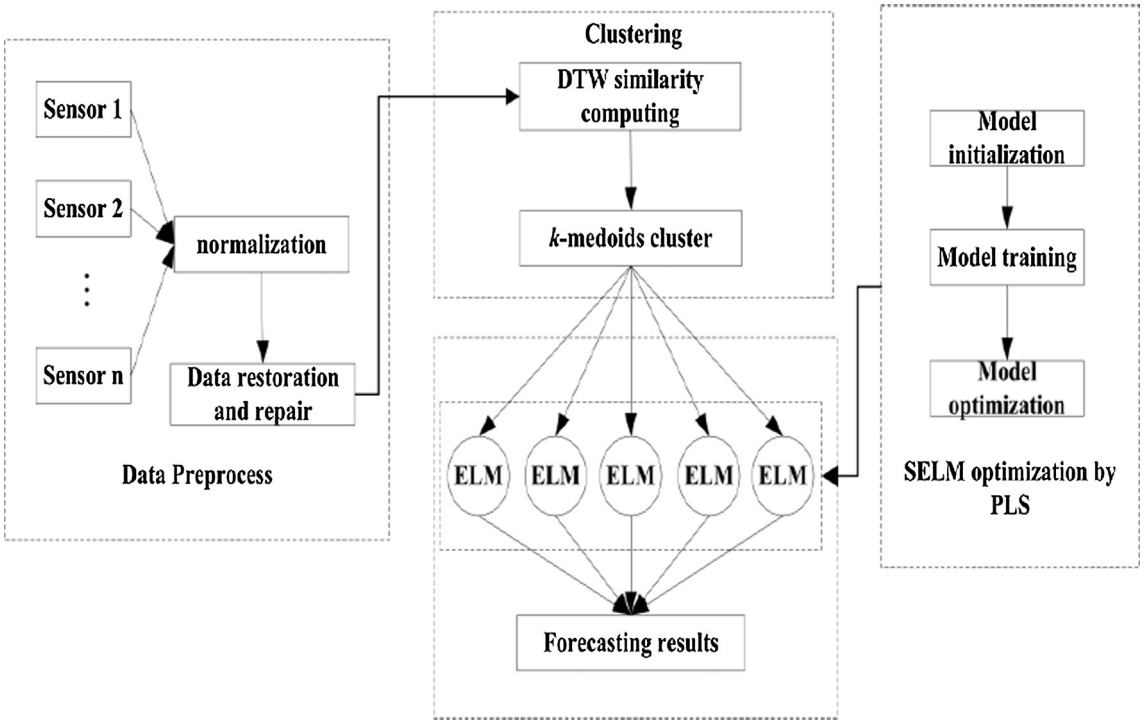


Fig. 3. The detail process of dissolved oxygen content forecasting.

verify whether the calculation of meteorological index is reliable. Moreover, we used linear interpolation method (Silva et al., 2003) to restore missing data. And "bad data" is corrected by selecting a data with the smallest difference to replace the erroneous data, and thus ensured that the data fluctuate less than 10%. We then focus on presenting the detailed techniques of Steps 2–3 in the following subsections.

2.3. The k-medoids clustering based on segmentation

Clustering based on reasonable time segmentation ensures the efficiency of the prediction. Clustering plays a crucial role in improving the efficiency of the CSELM model. Time segmentation can help to find the periodicity of time series and thus increase the prediction performance.

2.3.1. Segmentation

Given the similar trend existed in cycle time, we subdivided the data samples into day-stream and night-stream. It is easy to do so. Because the water quality data is collected in summer in this paper, the sunrise always occur around 6:00 am, and sunset occurs at about 7:00 pm. Therefore we set the day data set from 6:00 am to 18:59 pm and night data set from 7:00 pm to 5:59 am.

Generally the day is longer than night, and the start and end time of data collection are not at the sunrise and sunset, the lengths of day and night data sets that need to be analyzed are different. DTW distance can be effectively and efficiently employed for similarity calculation (Cao et al., 2016). Because time series data sets vary in length, we utilize the DTW distance to finish calculation of time series after data preprocess.

To explain how to calculate time series similarity, we illustrate the calculation process with an example that involves two time series: $R = \{R_1,R_2,\ldots,R_m\}$ of length m and $T = \{T_1,T_2,\ldots,T_n\}$ of length n, where m may not equate to n. We present the process of using DTW distance to calculate the similarity of time series as follows. We establish the distance matrix $D_{m\times n}$ whose element $d_{ij} = (R_i\text{-}T_j)^2$. The similarity value S can be calculated as follows:

$$S(R, T) = Dist = \min\left\{\frac{1}{k}\sqrt{\sum_{i-1}^{k} w_k}\right\} \tag{1}$$

where $Dist$ denotes the DTW distance between R and T, $W = \{w_1,w_2,\ldots,w_k\}$ denotes the warping path, which is a series of neighboring elements in the distance matrix and achieves the least cumulative d_{ij} values along the path (Long et al., 2012).

Boundary condition, continuity condition and monotonicity condition compose the main constraints on the warping path (Berndt and Clihord, 1994). Boundary refers to the warping paths from $(i_1, j_1) = (1, 1)$ to $(i_k, j_k) = (m, n)$. Continuity denotes the steps in the matrix confined to the points with $i_k\text{-}i_{k-1} \le 1$ and $j_k\text{-}j_{k-1} \le 1$. Monotonicity means that the values of m and n of warping path increase monotonically.

To search the best warping distance, dynamic programming is used to construct $Dist$. The cumulative distance matrix is defined as follows.

$$Dist(i, j) = d_{ij} + \min(Dist(i-1, j), Dist(i, j-1), Dist(i-1, j-1)) \tag{2}$$

Table 1
The CSELM prediction model.

Algorithm 1 CSELM Prediction modeling approach	
Step 1	Data preprocessing. After data normalization, finish the processes of repairing anomalous data and restoring missing data
Step 2	*k*-medoids clustering based on DTW time series similarity. Calculate the similarity of time slots based on DTW distance, and finish clustering the data sets based on similarity
Step 3	CSELM modeling. After clustering, train and build prediction models with CSELM algorithms in each cluster. And then test the built models using different indicators in testing data sets to finish forecasting

where, $Dist(0,0) = 0$, $Dist(i,0) = Dist(0,j) = +\infty$ and $Dist(i, j)$ is the sum of current d_{ij} and the minimum of the cumulative distance of previous elements.

2.3.2. Clustering

The k-medoids was applied for clustering of the water quality time series based on time segmentations. Compared with the k-means, k-medoids has a different mechanism for updating center location of a cluster. In k-means, the center of a cluster is determined by the mean location of all the data objects in the cluster, and thus the cluster center is a virtual data object. In k-medoids, however, it chooses a real data object as the cluster center which has the minimum sum of distance to other data objects in the cluster. k-medoids is more reliable and robust than k-means to hand the outliers (Park and Jun, 2009).

The main idea of k-medoids is presented as follows. First, we set a value to k, the number of clusters. Second, we select k samples from all the data as the initial centers of k clusters. Third, we assign each building sample to the nearest cluster center based on a distance measure. Fourth, we find the median data object as a new cluster centre, i.e., the data object with the minimum average distance to the remaining data objects. Repeat the last two steps until there is no change in the center of clusters.

In this paper, the initial parameter of k was determined by our experience on extensive experimental study. We used the cluster validity index to assess the cluster performance. As the most common cluster validity index, Davies-Bouldin (DB) index is a function elaborating compactness in the same class and dispersion in different clusters (Sassi, 2012), and it can be defined as:

$$V_{DB}(k) = \frac{\sum_{i=1}^{k} \max_{j,j\neq i}\left\{\frac{S_i + S_j}{d_{ij}}\right\}}{k} \tag{3}$$

where $S_i = \frac{1}{n_i}\sum_{x\in C_i}\|x - z_i\|$ is compactness in cluster C_i, $d_{ij} = \|z_i - z_j\|$represents the dispersion between cluster C_i and cluster C_j. The smaller DB value represents the better clustering result. In this paper, we set the number of cluster, k, between 2 and 6. The DB values for different k are given in Table 2.

In Table 2, $DB = 0.8384$ is the smallest value, so $k = 5$ is the best; and thus the 63 sets of dissolved oxygen time series are divided into 5 clusters. The clustering results are shown in Fig. 4. From Fig. 4, it can be seen that the data objects in each cluster are concentrated and there is relatively high dispersion between clusters.

2.4. PLS Softplus ELM (PLS-SELM)

We propose a new PLS-SELM method, which is a multi-input, single output structure with activation function of Softplus based on PLS. k ELM sub-models are provided for k clusters, which utilize PLS algorithm to solve the strong co-linear problem and get output weight of ELM based on Softplus in each cluster. We can obtain k PLS-SELM models and compare the performance based on different indicators in various clusters.

2.4.1. The ELM algorithm

Extreme learning machine (ELM) was proposed by Huang et al. (2006) for single-hidden layer feed-forward neural network. ELM consists of three layers: input layer, hidden layer and output layer.

Given a data set $= (x_i, t_i)$, $i = 1,2,\ldots N$, where $x_i = [x_{i1},x_{i2},\ldots,x_{in}]^{\mathrm{T}}$ is the i^{th} sample, $t_i = [t_{i1},t_{i2},\ldots,t_{im}]^{\mathrm{T}}$ is the actual tag of the i^{th} sample, n is the dimension of each sample, and m is class number of total samples. The standard SLFNs with L hidden layer nodes and activation function $g(x)$ are mathematically modeled as follows:

$$\sum_{j=1}^{L} \beta_j g(w_j x_i + b_j) = t_i,\ i = 1, 2, \ldots, N \tag{4}$$

where $w_j = [w_{j1},w_{j2},\ldots,w_{jn}]^{\mathrm{T}}$ is the weight vector connecting the j^{th} hidden node and the input nodes, $\beta_j = [\beta_{j1},\beta_{j2},\ldots,\beta_{jm}]^T$ is the weight vector connecting the j^{th} hidden node and the output nodes, b_j represents the bias of the j^{th} hidden layer neuron, and $w_j x_i$ denotes the inner product of w_j and x_i. Eq. (4) can be expressed in an explicit matrix form:

$$H\beta = T \tag{5}$$

In the Eq. (5),

$$H(w_1,\cdots,w_L, b_1,\cdots,b_L, x_1,\cdots,x_N) = \begin{bmatrix} g(w_1x_1 + b_1) & \cdots & g(w_Lx_1 + b_L) \\ \vdots & & \vdots \\ g(w_1x_N + b_1) & \cdots & g(w_Lx_N + b_L) \end{bmatrix}$$

and $\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_L^{\mathrm{T}} \end{bmatrix}_{L\times m}$, $T = \begin{bmatrix} t_1^{\mathrm{T}} \\ \vdots \\ t_N^{\mathrm{T}} \end{bmatrix}_{N\times m}$ where H represents the output matrix of hidden layer, the j^{th} column of H is the output of j^{th} hidden layer unit, $j = 1,2,\ldots, L$.

The training of ELM is a process of solving linear least squares between the hidden layer and output layer. The output weight $\hat{\beta}$ is the unique tunable parameters, the least-square solution can be written as:

$$\hat{\beta} = H^{+}T \tag{6}$$

where H^{+} denotes the Moore-Penrose inverse of hidden layer output matrix H, $H^{+} = (H^TH)^{-1}H^T$.

2.4.2. Softplus ELM

The activation function has significant influence on forecasting accuracy in the ELM network training. Traditional Sigmoid is a good threshold function, and it is often used as the activation function in ELM. The function is defined as follows:

$$g(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

Actually, the activation function is not unique (Hornik, 1991). Rectified Linear Unit Function (ReLU) is a new type of function in deep learning field, and it can be defined as follows:

$$g(x) = \max(0, x) \tag{8}$$

Compared to the Sigmoid, ReLU is closer to the activation model of biology, and it has the advantages of high efficiency and effective - generalization for its simplicity. Nevertheless, the sparse ability of ReLU is realized by forcing some data to be zero, which will weaken the forecasting ability. ReLu is a piecewise function, and the derivative of ReLu is discontinuous. Softplus is an analytic form of smoothing approximation to ReLU presented by Glorot et al. (2011), and it is defined as follows:

$$g(x) = \ln(1 + e^{x}) \tag{9}$$

The Softplus is a nonlinear, continuous and differentiable function and closer to the activation model of biology than Sigmoid. In this paper, for the effective generalization, we introduce the Softplus to the ELM to replace the Sigmoid called Softplus improved ELM.

2.4.3. PLS-SELM

Hidden output matrix will be strong co-linear if the input nodes are redundant or the number of hidden layer is more than samples. These can cause that the inverse matrix of H^TH doesn't exist and the least-square solution of output weight $\hat{\beta}$ is unstable. To avoid these problems

Table 2
DB values change with k.

Cluster number k	2	3	4	5	6
DB	1.2310	0.8774	0.9420	0.8384	0.8945

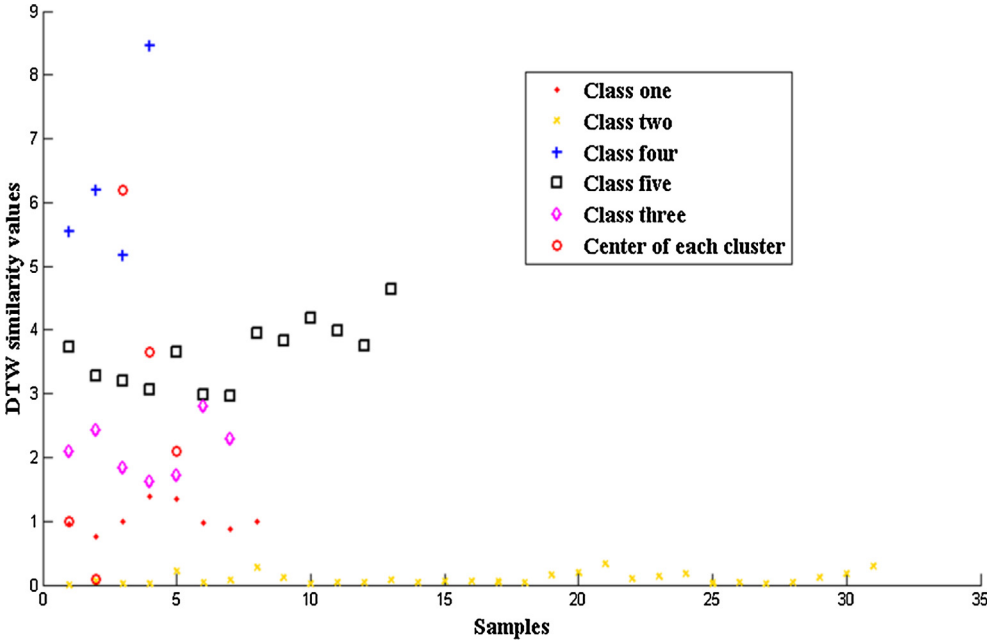


Fig. 4. The clustering results.

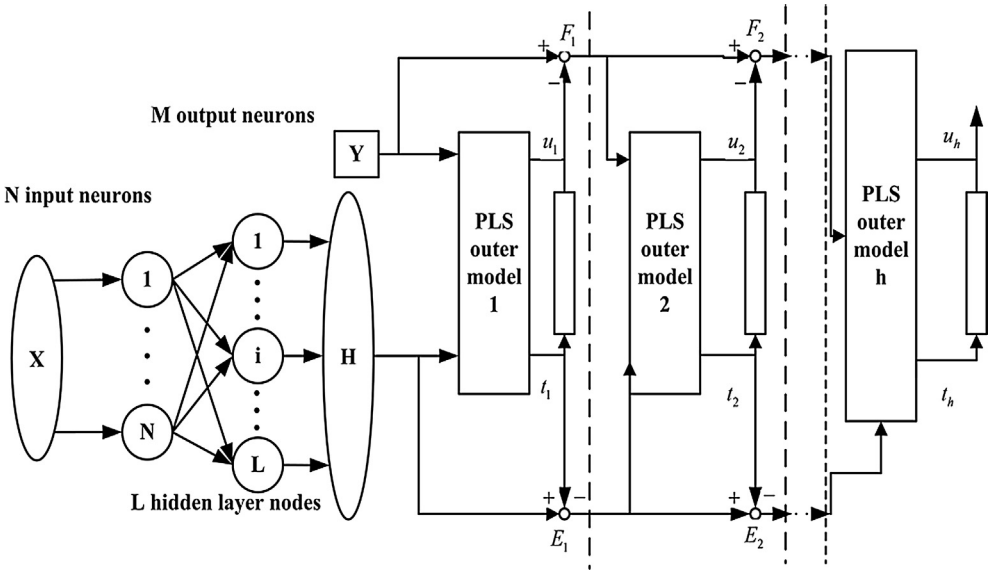


Fig. 5. The Schematic diagram of PLS-SELM.

in ELM training, PLS was embedded into the Softplus ELM model, and PLS will replace the least-square method to solve the strong co-linear of hidden output matrix through extracting the orthogonal variable in hidden layer. The schematic diagram of PLS-SELM is shown in Fig. 5.

PLS can build the linear relation between the output $Y_{N\times M}$ and hidden layer nodes output matrix $H_{N\times L}$ during PLS-IELM modeling as follows.

$$Y = H\beta_{PLS} + e \tag{10}$$

where β_{PLS} and e are the output weight coefficient matrix and noise in the PLS-SELM model respectively. The bilinear decomposition between output matrix H of hidden layer nodes and output matrix Y can be described as follows:

$$\left.\begin{aligned} H &= TP^{\mathrm{T}} + E = \sum_{k=1}^{h} t_k p_k^{\mathrm{T}} + E, \\ Y &= UQ^{\mathrm{T}} + F = \sum_{k=1}^{h} u_k q_k^{\mathrm{T}} + F. \end{aligned}\right\} \tag{11}$$

where $T = [t_1, \cdots, t_h] \in R^{N\times h}$ and $U = [u_1, \cdots, u_h] \in R^{N\times h}$ denote the score matrixes of hidden layer and output layer respectively. $P = [p_1, \cdots, p_h] \in R^{L\times h}$ and $Q = [q_1, \cdots, q_h] \in R^{m\times h}$ denote the load matrixes of hidden layer and output layer respectively. h is the number of latent variables retained in the PLS model. $E_{N\times L}$ and $F_{N\times m}$ are residual error matrixes of hidden layer and output layer. The internal model between hidden layer and output layer is a linear regression model. It is established based on each latent variables u_k and t_k as follows:

$$u_k = t_k b_k;\ k = 1,\ \cdots, h. \tag{12}$$

where $b_k = (t_k^{T} t_k)^{-1} t_k^{T} u_k$, b_k is the least squares coefficient of latent variables. $B = diag[b_k]$, the diagonal element of B is scalar b_k. The internal model can be respresented in matrix form as:

$$U = TB \tag{13}$$

Then, the parameters of model between hidden layer and output layer can be solved by nonlinear iterative PLS algarithm (NIPALS). The

Table 3
Neural network structures in five clusters.

Serial number of clusters	Total data points	Training data points	Testing data points	Structure of neural network	RMSE
1	528	528	0	–	–
2	2418	2028	390	11–50-1	0.466
3	462	396	66	11–22-1	0.3294
4	228	198	30	11–16-1	0.2493
5	828	630	198	11–29-1	0.335

input-output relation between hidden layer and output layer can be described as$\{H, Y\} \overset{PLS}{\rightarrow} \{T, W, P, B, Q\}$, where W is the input weights. If the first h latent variables are retained, the output weights $\hat{\beta}_{PLS}$ in PLS-SELM model can be calculated as:

$$\hat{\beta}_{PLS} = W(P^{T}W)^{-1}BQ^{T} \tag{14}$$

For each test time series data, we find out the cluster it belongs to and then compute the forecasting results by Eq. (10).

3. Results and discussions

3.1. Parameter settings of algorithm

In this experiment, all data (4464 data sets) are grouped into 31 streams (each for a day). We segment each stream into two sub pieces: day-stream and night-stream, since the monitoring data are changed periodically in day and night; so, totally 63 time series data sets are used. In ELM, the numbers of hidden nodes L in different clusters are determined by using trial and error method. And the number of the hidden nodes in the five clusters are $L_2 = 50$, $L_3 = 22$, $L_4 = 16$, $L_5 = 29$ respectively, as shown in Table 3. In this experiment, the partial least squares algorithm is utilized to get output weight of ELM. In PLS, the number of latent variables h also need to be selected by using the cross-validation method and the performance indictor of Root Mean Square Error (RMSE). The minimum RMSE value occurred at latent variables $h = 5$. All of the experiments are implemented by MATLAB and run on a PC with 3.4 GHz Core (TM) processor, 4.0G memory, and Microsoft Windows 7.

3.2. Results of k-medoids clustering

As shown in Table 2, the optimal number of clusters is 5 for the 63 data sets. Each cluster contains a CSELM prediction sub-model. Table 3 compares the neural network structures and RMSE values of the five clusters. Each cluster has different structure. It shows that no testing data point is assigned to the first cluster. And the biggest cluster contains 2418 data points and the smallest cluster contains the only 228 data points. The second cluster has 50 (the greatest number of) hidden layers with the structure of 11-50-1. We can find that all the data points in the second cluster are night-streams, and day-streams are clustered into the other four clusters. So, the difference of data points in day is greater than at night. In view of RMSE value in these clusters, the greater the total data point, the greater the RMSE value.

3.3. Model performance evaluation

3.3.1. Evaluation metrics

We evaluate the prediction accuracy using the four metrics: RMSE, Mean Absolute Percentage Error (MAPE), Nash Sutcliffe efficiency Coefficient (NSC) (Benyahya et al., 2007) and Mean Absolute Error (MAE), and also evaluate the Run Time T. These metrics are computed by Eqs. (15)–(18).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{15}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{16}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{17}$$

$$NSC = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{18}$$

where N is the total number of actual points in each data set, y_i is the original and $\hat{y}_i$ is the prediction value. $\bar{y}$ and $\tilde{y}$ represent the average of observed value and the average of prediction values respectively. The higher NSC value, shorter T value, lower MAPE and MAE values indicate more precise model.

3.3.2. The best proposed method

We evaluate the two critical components (clustering and Softplus) of CSELM by adding them to improve ELM and PLS-ELM separately; so we propose two more new methods: SELM (improving ELM by Softplus and PLS) and CELM (improving PLS-ELM by clustering input).

Fig. 6 shows the average precision and time of prediction models. The overall trend is that CSELM exceeds both SELM and CELM in terms of precision (RMSE, MAPE and MAE) and time, though the NSC value of CSELM is only slightly higher than that of SELM but obviously higher than that of CELM. We can see that the average run time of CSELM is far shorter than SELM; this demonstrates that clustering of raw data using DTW greatly improve the efficiency. We also can see that the average run time of CSELM is shorter than CELM; this means that Softplus also contribute to the efficiency improvement. However, CELM achieves worse precision (RMSE, MAPE and MAE); this demonstrate that Softplus in both CSELM and SELM improves the accuracy. Moreover, the best NSC value of CSELM among these prediction methods indicates that the combination of the clustering operation, PLS and Softplus can improve the performance of dissolved oxygen prediction well.

3.3.3. Comparison with existing methods

For further analyze the applicability of CSELM, we use the trained models to predict the dissolved oxygen content in five days and compare CSELM with two counterpart methods (PLS-ELM and ELM) as shown in Fig. 7 (a-e). In these figures, the x coordinate denotes the time in a day and y coordinate represents the dissolved oxygen content.

We can find a common trend in Fig. 7 (a-e) that all models can complete the dissolved oxygen prediction well, but vary in prediction effect. The forecasting curve of CSELM is closer to the real data than those curves of PLS-ELM and ELM. This demonstrates that CSELM

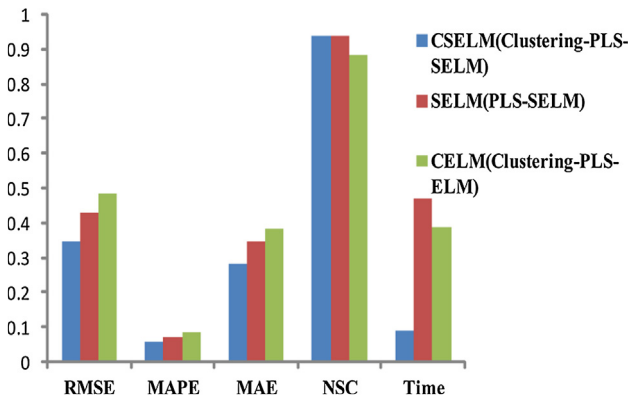
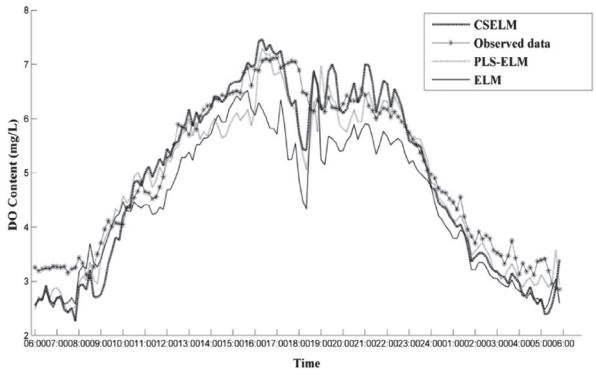
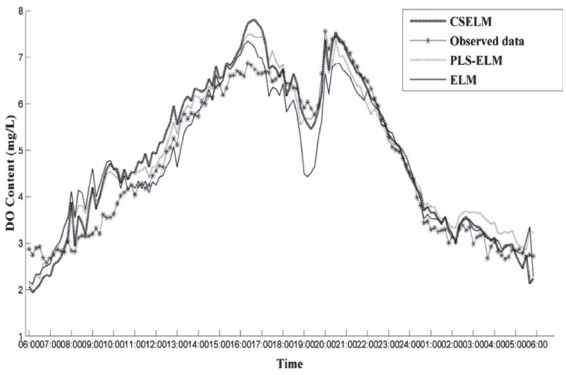


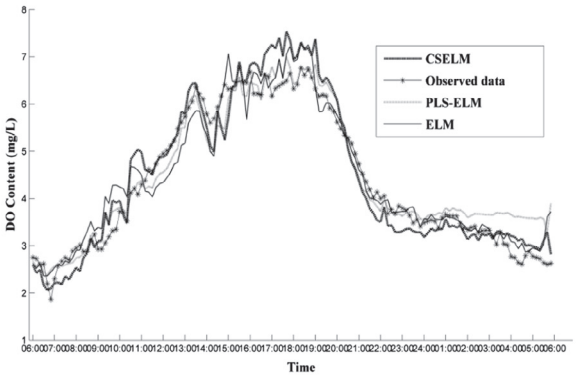
Fig. 6. Average precision and time.



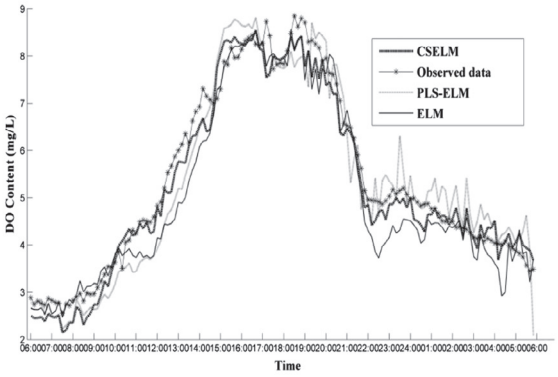
(a) The forecasting results on the first day



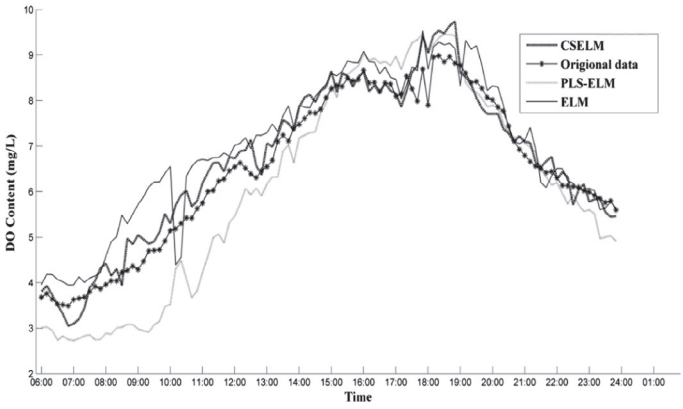
(b) The forecasting results on the second day



(c) The forecasting results on the third day



(d) The forecasting results on the fourth day



(e) The forecasting results on the fifth day.

Fig. 7. The forecasting results of dissolved oxygen content by CSELM.

achieves the best forecasting results by recognizing the periodic change patterns and trend changes of time series data. However, the predicting results of dissolved oxygen content changed sharply on the peak of the first day in Fig. 7 (a), on the peak of the second day in Fig. 7 (b) and the valley of the fifth day in Fig. 7 (e). Through careful observation and analysis, we have found that there are some fluctuations in the Fig. 7 (a-e) occurred in half an hour before and after 6o'clock am or 7o'clock pm. The two points are not just the sunrise and sunset, but also the critical points to divide the similar time slots in Section 2.3. Due to the time differences of every sunrise and sunset, the start time and end time of photosynthetic algae are different. Thus, the forecasting based on similar time slot had small errors when dissolved oxygen changed sharply.

We also provide some numeric predicted results of examples in Table 4 during twenty minutes before and after 7 pm on the first day and twenty minutes before and after 6 am on the fifth day. CSELM produced lower average absolute percentage error (APE) than PLS-ELM and ELM on the first and the fifth days. The average APE values for CSELM, PLS-ELM and ELM were 0.3294, 1.3365 and 0.7295 during twenty minutes before and after 7 pm on the first day. Meanwhile, the average APE values for CSELM, PLS-ELM and ELM modeling approaches were 0.2034, 0.5649 and 1.0718 during twenty minutes before and after 6 am on the fifth day. It is obviously that, there are big fluctuations in these positions, but CSELM model can achieve better forecasting results than PLS-ELM and ELM modeling approaches.

In addition to the best prediction effectiveness of CSELM as shown in Fig. 7, we give the numeric precision (RMSE, MAPE, MAE, NSC) values in Table 5. The average RMSE of the proposed CSELM is the best

Table 4
Comparison of forecasting results.

Time (am)	RV (mg/L)	CSELM		PLS-ELM		ELM	
		PV (mg/L)	APE	PV (mg/L)	APE	PV (mg/L)	APE
18:40	6.1306	6.8599	0.7293	5.0720	1.058565	5.2633	0.867282
18:50	6.3384	6.6127	0.2743	4.7356	1.602773	5.3193	1.019121
19:00	6.2306	6.1530	0.0776	4.9793	1.251269	5.6714	0.559217
19:10	6.1228	6.2163	0.0935	4.7486	1.374166	5.4780	0.644752
19:20	6.3888	6.8610	0.4722	4.9931	1.395722	5.8316	0.557197
5:40	3.4498	3.8890	0.4392	4.1001	0.650273	4.5198	1.07002
5:50	3.4860	3.6726	0.1866	4.1106	0.624648	4.5549	1.06886
6:00	3.6781	3.8024	0.1243	4.1458	0.467704	4.591	0.912945
6:10	3.7616	3.9338	0.1722	4.2708	0.509227	4.8918	1.130247
6:20	3.6403	3.7350	0.0947	4.2133	0.573008	4.8171	1.176754

*RV means the real value; PV means the predictive value; APE means the absolute percentage error.

Table 5
Precision statistics on three models.

Index	CSELM	PLS-ELM	ELM
RMSE	0.3449	0.681	0.7573
MAPE	0.058	0.1045	0.1287
MAE	0.2806	0.5298	0.6185
NSC	0.9392	0.8453	0.8087
Time	0.3684	0.9975	0.7829

(0.3449), compared to PLS-ELM (0.681) and ELM (0.7573). Also, the average MAPE of the proposed CSELM is improved 44.5% compared to PLS-ELM and improved 54.93% compared to ELM. Similarly, the average MAE value of our CSELM is improved 47.03% of PLS-ELM and 54.62% of ELM. It is obvious that the proposed CSELM model outperforms the two counterparts in terms of all precision metrics for dissolved oxygen prediction. Meanwhile the CSELM model produced better NSC value (0.9392) in comparison to the two counterparts with values of 0.8453 and 0.8087. These data demonstrate the good performance of CSELM prediction model. CSELM also significantly increases the efficiency, compared to both PLS-ELM and ELM. The run time of ELM is nearly double of CSELM, while the run time of PLS-ELM is almost three times of CSELM.

In summary, our proposed CSELM is obviously the best model for predicting the water quality changes in terms of accuracy and efficiency, compared to the two counterparts (PLS-ELM and ELM). Particularly, CSELM is able to discover the periodic change patterns and trend changes of water quality and meteorological time series data (in aquaculture) and thus provide better prediction results.

4. Conclusions

We proposed prediction model CSELM for forecasting dissolved oxygen content change from time series data, which comprises two new techniques: the *k*-medoids clustering based on DTW for both accuracy and efficiency by intelligently grouping input data based on their common trends, and Softplus activation function based on PLS for improving ELM. The experimental study shows that CSELM performs better than two counterparts (PLS-ELM and ELM) models in terms of RMSE, MAPE, MAE, NSC, and Time. CSELM provided a new forecasting method for various-length time slots of day or night and its effectiveness was verified in a real-world application for predicting dissolved oxygen to reduce and avoid unnecessary losses caused by hypoxia.

Acknowledgment

This study was supported in part by the National Natural Science Foundation of China (Grant No. 61472368), Central Public-interest Scientific Institution Basal Research Fund, China (Grant No. 2016HY-ZD1404), Key Research and Development Project of Jiangsu Province, China (Grant No. BE2016627), the Fundamental Research Funds for the Central Universities, China (Grant No. RP51635B), and Wuxi International Science and Technology Research and Development Cooperative Project, China (Grant No. CZE02H1706).

Conflicts of interest

None.

References

Amid, S., Gundoshmian, T.M., 2017. Prediction of output energies for broiler production using linear regression, ann (mlp, rbf), and anfis models. Environ. Prog. Sustainable Energy 36 (2), 577–585. https://doi.org/10.1002/ep.12448.

Antanasijević, D., Pocajt, V., Povrenović, D., Perićgrujić, A., Ristić, M., 2013. Modeling of dissolved oxygen content using artificial neural networks: danube river, north serbia, case study. Environ. Sci. Pollut. Res. 20 (12), 9006–9013. https://doi.org/10.1007/s11356-013-1876-6.

Bartzanas, T., Bochtis, D.D., Green, O., Sørensen, C.G., Fidaros, D., 2013. Prediction of quality parameters for biomass silage: a cfd approach. Comput. Electron. Agric. 93 (2), 209–216. https://doi.org/10.1016/j.compag.2012.05.013.

Benyahya, L., Hilaire, A.S., Quarda, B.M.J.T., Bobee, B., Nedushan, B.A., 2007. Modeling of water temperatures based on stochastic approaches: case study of the Deschutes River. J. Environ. Eng. Sci. 6, 437–448. https://doi.org/10.1139/S06-067.

Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: Workshop on Knowledge Discovery in Databases, vol. 10, pp. 359–370.

Cao, Y., Rakhilin, N., Gordon, P.H., Shen, X., Kan, E.C., 2016. A real-time spike classification method based on dynamic time warping for extracellular enteric neural recording with large waveform variability. J. Neurosci. Methods 261, 97–109. https://doi.org/10.1016/j.jneumeth.2015.12.006.

Chen, W., Yuan, H.M., 2014. An improved GA-SVM algorithm. In: Proc. the Conference on Industrial Electronics and Applications, pp. 2137–2141.

Faruk, D.Ő., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586–594. https://doi.org/10.1016/j.engappai.2009.09.015.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. Proceeding of the Conference on Artificial Intelligence and Statistics.

Heddam, S., Kisi, O., 2017. Extreme learning machines: a new approach for modeling dissolved oxygen (do) concentration with and without water quality variables as predictors. Environ. Sci. Pollut. Res. Int. 24, 16702–16724. https://doi.org/10.1007/s11356-017-9283-z.

Hatzikos, E., Anastasakis, L., Bassiliades, N., Vlahavas, I., 2005. Simultaneous prediction of multiple chemical parameters of river water quality with tide. Proceedings of the 2nd International Scientific Conference on Computer Science. IEEE Computer Society, Bulgarian Section.

Hornik, Kurt, 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.

Huan, J., Cao, W.J., Liu, X.Q., 2017. A dissolved oxygen prediction method based on k-means clustering and the elm neural network: a case study of the changdang lake, china. Appl. Eng. Agric. 33 (4), 461–469. https://doi.org/10.13031/aea.11786.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1–3), 489–501.

Junior, J.J.D.M.S., Backes, A.R., 2016. Elm based signature for texture classification. Pattern Recogn 51 (C), 395–401. https://doi.org/10.1016/j.patcog.2015.09.014.

Liu, S., Xu, L., Li, D., Li, Q., Jiang, Y., Tai, H., et al., 2013. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. Comput. Electron. Agric. 95 (4), 82–91. https://doi.org/10.1016/j.compag.2013.03.009.

Liu, S., Xu, L., Li, D., Zeng, L., 2014. Online prediction for dissolved oxygen of water

quality based on support vector machine with time series similar data. Trans. Chinese Soc. Agric. Eng. 30 (3), 155–162.
Long, X., Fonseca, P., Foussier, J., Haakma, R., 2012. Using dynamic time warping for sleep and wake discrimination. In: Proceedings of the IEEE-Embs International Conference on Biomedical and Health Informatics, pp. 886–889.
Mahapatra, S.S., Nanda, S.K., Panigrah, B.K., 2011. A cascaded fuzzy inference system for Indian river water quality prediction. Adv. Eng. Softw. 42 (10), 787–796. https://doi.org/10.1016/j.advengsoft.2011.05.018.
Palani, S., Liong, S.Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Mar. Pollut. Bull. 56 (9), 1586–1597.
Park, H.-S., Jun, C.-H., 2009. A simple and fast algorithm for *k*-medoids clustering. Expert Syst. Appl. 36 (2), 3336–3341.
Sassi, M., 2012. Towards fuzzy-hard clustering mapping processes. Comput. Sci. 1, 37–63.
Shahin, M.A., 2016. State-of-the-art review of some artificial intelligence applications in pile foundations. Geosci. Front. 7 (1), 33–44. https://doi.org/10.1016/j.gsf.2014.10.002.
Silva, M.C.G., Pereira, C.A.C., Cruz, J.M.S., 2003. On the use of a linear interpolation method in the measurement procedure of a seven-hole pressure probe. Exp. Therm. Fluid Sci. 28 (1), 1–8. https://doi.org/10.1016/S0894-1777(03)00074-8.
Tan, G., Yan, J., Gao, C., Yang, S., 2012. Prediction of water quality time series data based on least squares support vector machine. Procedia Eng. 31 (16), 1194–1199.
Xu, L., Liu, S., 2013. Study of short-term water quality prediction model based on wavelet neural network. Math. Comput. Model. 58 (3–4), 807–813. https://doi.org/10.1016/j.mcm.2012.12.023.
Yan, K., Ji, Z., Lu, H., Huang, J., Shen, W., Xue, Y., 2017. Fast and accurate classification of time series data using extended elm: application in fault diagnosis of air handling units. IEEE Trans. Syst. Man Cybernet. Syst. PP (99), 1–8.
Yan, K., Zhong, C.W., Ji, Z.W., Huang, J., 2018. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy Build. 181, 75–83.
Ye, X., Shi, P., Kuang, L., 2015. Research on comprehensive ability of Zhejiang Province in typhoon disaster resistance and reduction. J. Nat. Disasters 1, 214–221.
Yu, H., Chen, Y., Hassan, S., Li, D., 2016. Dissolved oxygen content prediction in crab culture using a hybrid intelligent method. Sci. Rep. 6, 27292. https://doi.org/10.1038/srep27292.
Zhao, L., Wang, D., Chai, T., 2013. Estimation of effluent quality using pls-based extreme learning machines. Neural Comput. Appl. 22 (3–4), 509–519. https://doi.org/10.1007/s00521- 012-0837-1.
Zhou, J., Liu, Y., Zhang, T.H., 2016. Analytical redundancy design for aeroengine sensor fault diagnostics based on SROS-ELM. Math. Probl. Eng. 9. https://doi.org/10.1155/2016/8153282. Article ID. 8153282.
Zhu, X., Li, D., He, D., Wang, J., Ma, D., Li, F., 2010. A remote wireless system for water quality online monitoring in intensive fish culture. Comput. Electron. Agric. 71 (1), S3–S9. https://doi.org/10.1016/j.compag.2009.10.004.