



Prediction model of pH value in mitten crab culture

CHENGYUN ZHU*, XINGQIAO LIU, HAILEI CHEN AND XIANG TIAN

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang - 212 013, China

*School of New Energy and Electronic Engineering, Yancheng Teachers University, Yancheng - 224 005, China

e-mail: xqliu@ujs.edu.cn

ABSTRACT

The pH of water directly affects growth of mitten crab (*Eriocheir sinensis* H. Milne-Edwards, 1853) in aquaculture. A prediction model was set up to determine the changing trend of pH value during culture of mitten crabs. The model would help the farmer to take measures in advance to maintain the safety of cultured crabs, when the predicted value of pH is found to cross beyond safe levels. Prediction model of pH is based on the least squares support vector regression (LSSVR) model with chaotic mutation to improve the estimation of the distribution algorithm (CMEDA) to find optimal parameters (γ and σ) of LSSVR. Because these two parameters can significantly affect the performance of the LSSVR, the other three parameter optimisation methods viz., the particle swarm optimisation (PSO) algorithm, the genetic algorithm (GA) and grid search (GS) algorithm were used to compare with the CMEDA algorithm. The calculated mean absolute percentage errors of the results of the four prediction models were 0.4059, 0.6332, 0.9385 and 1.2499%, respectively. The CMEDA-LSSVR model has a higher prediction accuracy and more reliable performance than the other models. The prediction model was used in Xinhua, Jiangsu Province, China and it performed well and helped farmers make decisions and reduce aquaculture risks.

Keywords: Distribution algorithm, Least-squares support vector machines, Optimisation, pH value, Prediction model

Introduction

Freshwater shellfishes are sensitive to changes in pH of water. The change in water pH can affect the survival rate and growth of the animals as well as mineralisation of shells (Cai *et al.*, 2012). pH value is an important water quality parameter in mitten crab culture (Ritvo *et al.*, 2003) and the suitable pH range is generally considered between 7.5 and 8.5 (Hong *et al.*, 2013). pH value in aquatic environment fluctuates during day and night. Due to photosynthesis by phytoplankton, the pH value of culture ponds can reach over 10.2 during the day, which can go below 6 during the night because of the respiration of aquatic plants and animals. Too low or too high pH values can lead to large scale mortality of cultured mitten crabs (Page *et al.*, 2014). Therefore, it was found quite necessary to build a prediction model of pH of pond water (Xu and Xu, 2016), which could provide a basis to the farmers' for decision-making and thus optimise their operations as well as reduce economic losses. Many factors could lead to the change of pH value of rearing water in crab culture and hence prediction model of pH value is an open, dynamic and nonlinear system. Researchers have studied many algorithms to establish the prediction model of pH such as, nonlinear model (Ekici *et al.*, 2014), adaptive neuro-fuzzy inference system (ANFIS) model (Gaya *et al.*, 2014), adaptive genetic hybrid neural network (AG-HNN)

(Tang *et al.*, 2013) and grey model (Li and Cai, 2013). But these prediction models have various disadvantages, such as over-fitting, poor generalisation capability, locally optimal solutions, poor stability and so on.

Support vector machine (SVM) was first proposed by Vapnik (Vapnik, 1995) and its excellent performance in regression problems even for small samples has been demonstrated. The least squares support vector machine (LSSVM) (Stephen *et al.*, 2014), a least squares version of SVM, was proposed to overcome the shortcoming of high computational cost. LSSVM can be divided into LSSVR which is used for regression purposes and LSSVC for classification purposes. Some researchers have established pH prediction model using LSSVR (Goodarzi *et al.*, 2010; Dong *et al.*, 2014). The performance of LSSVR is heavily influenced by the two parameters (γ and σ), which are defined as an optimisation problem. So far, there is no accurate method to find the optimal parameters of LSSVR, therefore, an optimisation algorithm must be applied to find the optimal parameters.

Many algorithms were proposed to find the optimal parameters of LSSVR, such as: particle swarm optimisation (Lou *et al.*, 2014, Zheng *et al.*, 2014), genetic algorithm (Wang and Wang, 2012), different evolution (Stephen *et al.*, 2014; Sun and Liang, 2014) and gradient algorithm (Chen *et al.*, 2014). All these

methods can predict pH value, but the prediction accuracy could be further improved by adopting different parameter optimisation algorithm. Therefore, in this study, we propose a prediction model based on LSSVR using an improved distribution algorithm to estimate the optimal parameters for more accurate prediction of pH value in aquaculture of mitten crab *Eriocheir sinensis* H. Milne-Edwards, 1853.

Materials and methods

Data collection

The crab culture pond selected for the study was located at Xinhua City, Jiangsu Province, China. A monitoring system based on GPRS was designed to monitor the water quality parameters real-time. The structure diagram of the system is shown in Fig. 1. The system was mainly composed of a water quality parameter measuring instrument, GPRS communication module, site monitoring and remote server monitoring. Measuring instrument included several water quality parameter sensors such as dissolved oxygen sensor, temperature sensor, pH sensor, turbidity sensor, salinity sensor and conditioning circuit.

Date recorded for a period of 30 days during 1st to 30th June 2015 was downloaded from the database of the monitoring system. The sampling interval was 10 min, and thus we obtained 4320 samples and the sampling data were distributed into two parts: 3600 samples of water quality data were used for the LSSVR model training and 720 samples were used as the test data to evaluate the performance of the LSSVR model.

Least squares support vector regression

The goal of LSSVR model is to build a function $y = f(x)$ between the vector of observations x and the desired y . In this study, the four input variables of the LSSVR model were: dissolved oxygen (DO), water

temperature (WT), turbidity (TB) and salinity (SAL). The output of the LSSVR model was pH. Therefore, $x=[DO, WT, TB, SAL]$ and $y=pH$. In feature space, LSSVR model take the form (Mahmoud, 2011):

$$f(x) = w^T \phi(x) + b \quad (1)$$

where $w \in R^N$ and $b \in R$, w denotes the adjustable weight vector, b is the bias; $\phi(\cdot)$ is a nonlinear mapping function which is used to map the input vectors into a higher dimensional feature space. In the feature space, the LSSVR can be transformed into the following optimisation problem (Igne *et al.*, 2010):

$$\min \theta = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \quad (2)$$

$$\text{s.t. } y_i = w^T \phi(x_i) + b + e_i, \quad i=1, 2, \dots, N \quad (3)$$

where γ is the regularisation parameter and e_i is the error variable. In order to solve this optimisation problem, the Lagrange function was formulated as follows (Alvarez Meza *et al.*, 2012):

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{y_i [w^T \phi(x_i) + b] - 1 + e_i\} \quad (4)$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ is the Lagrange multiplier vector. The Karush-Kuhn-Tucker (KKT) conditions were used for optimality by differentiating L in Eq. (4) with the variable w, b, e, α . After eliminating the variables w and e , the following linear set can be obtained.

$$\begin{bmatrix} 0 & \bar{1}^T \\ \bar{1} & \Omega + I/\gamma \end{bmatrix} \times \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (5)$$

where $y = [y_1, y_2, \dots, y_N]^T$, $\bar{1} = [1, 1, \dots, 1]^T$, $\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$, $K(x_i, x_j)$ is the kernel function on the input vectors $\{(x_i, x_j), i=1, 2, \dots, N; j=1, 2, \dots, N\}$. The commonly used kernel function is the Radial Basis Function (RBF) defined by (Lin *et al.*, 2013):

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \quad (6)$$

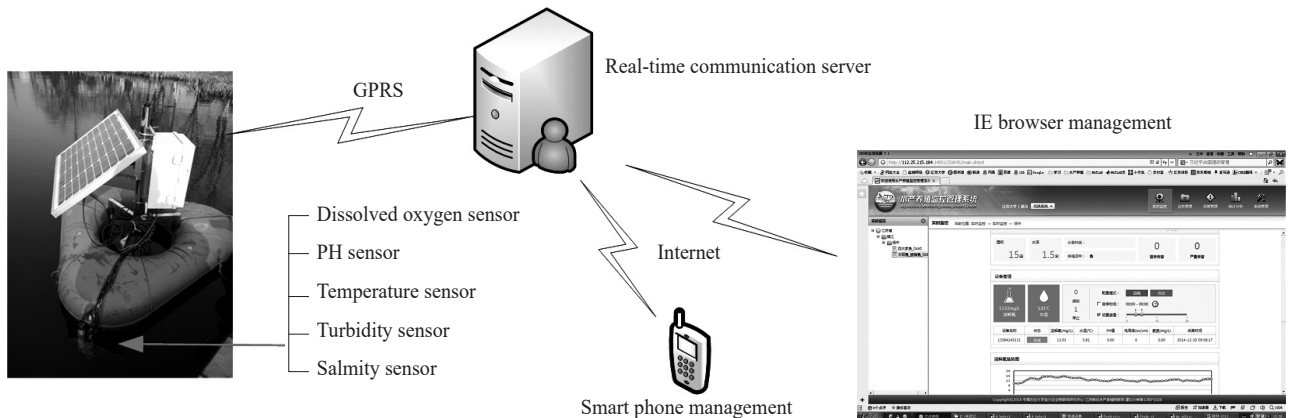


Fig. 1. Structure diagram of the real-time monitoring system

where σ is the width of radial basis function. Finally, the regression function can be represented in the following equation:

$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (7)$$

where α and b can be solved by Eq. (5). So LSSVR has two parameters: γ and σ , and since these two parameters can significantly affect LSSVR performance, the selection of these two parameters is the most important aspect. Some researchers adopted various methods to search for optimal parameters (Qiao, *et al.*, 2011; Liu *et al.*, 2014). In this study, estimation of distribution algorithms (EDA) was applied to find the optimal parameters γ and σ .

Estimation of Distribution Algorithm (EDA)

EDA has a good performance at solving complex problems in many areas. Although the performance of estimation of distribution algorithm is better than many optimisation algorithm, the estimation of distribution algorithm still has disadvantages. For instance, the individuals of the population are easy to trend to the same solution in the evolutionary process and the population diversity decreases sharply. These disadvantages affect the performance of the estimation of distribution algorithm. In order to keep population diversity, we used chaotic mutation to improve the estimation of distribution algorithm and obtained the CMEDA (Chaotic Mutation Estimation Distribution Algorithm) (Cheng *et al.*, 2010) and the CMEDA was used to optimise the two parameters of LSSVR.

Improved EDA

The chaotic motion can track any state in its scope without repetition according to its regularity. CMEDA is described in detail as follows:

CMEDA evolutionary process: Step 1. Since there were two LSSVR parameters to be optimised, two initial variables x_1, x_2 were generated randomly in the interval (0, 1). Let $X_0 = (x_1, x_2)$ be the initial node of one dimension logistic map model.

Step 1.1. One dimension logistic map was defined by:

$$X_{t+1} = \lambda \cdot X_t(1 - X_t) \quad (8)$$

where λ is control parameter. If given λ a value, after n times iteration, then generated n individuals $X_1, X_2, \dots, X_n$ by Eq. (8), the n individuals is a chaotic sequence. Let $X = \{X_1, X_2, \dots, X_n\}$ be a matrix with n rows and two columns, representing the initial population:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{n1} & X_{n2} \end{bmatrix} \quad (9)$$

Step 1. 2. The initial population X is expanded to the solution space and the i^{th} variable of the j^{th} individual can be represented in the following equation:

$$x_{ji} = \text{low} + (\text{high} - \text{low})x_{ji} \quad (10)$$

where the low is lower limits of the solution space and the high is upper limits.

Step 2. Fitness value for each individual of the current population was calculated. For individual X_j , the fitness value can be calculated as:

$$\text{fit}(X_j) = \frac{1}{n} \sum_j^n \sum_i^2 (d_{ji} - y_{ji})^2 \quad (11)$$

where d_{ji} is the i^{th} actual output value of the j^{th} individual, y_{ji} is the i^{th} predicted output value of the j^{th} individual.

Step 3. The new individuals were selected based on the mutation probability $p_m = 1$ to execute chaotic mutation. The mutation radius of i^{th} variable x_{ji} of the individual X_j was calculated according to the following equation:

$$r_{ji} = S(i) \frac{1 - \text{fit}(j)}{\max_{(1 \leq j' \leq n)} \text{fit}(j')} \quad (12)$$

where $S(i) = (2\max x_{*i} - \min x_{*i}) / n$ is mutation step-size. After chaotic mutating each new individual, the fitness value of these new individuals were calculated. New individuals are to be used to replace the old individuals when the fitness value of new individual is less than the fitness value of old individual and then obtain the new population, and go to Step 4. Otherwise keep the population without change and then go to Step 5.

Step 4. The output current is the best individual if it meets the stop condition of the algorithm, otherwise go to Step 5.

Step 5. Generate next population.

Step 5.1. Select np individuals from the present population. Then build a Gaussian model for all new individuals after statistical analysis of all new individuals.

Step 5.2. Generate $k \cdot np$ ($k > 1$) new individuals based on the Gaussian model.

Step 5.3. Calculate fitness values of these $k \cdot np$ new individuals, compare their fitness values and then choose the best individuals which are used as the next generation population.

Step 6. The output current is the best individual if it meets the stop condition of the algorithm, otherwise go to Step 3.

CMEDA-LSSVR model

Through the CMEDA evolutionary process, the two LSSVR parameters γ and σ were optimised, which build optimised LSSVR model for predicting pH value accurately. The procedure of CMEDA-LSSVR model is illustrated in Fig. 2.

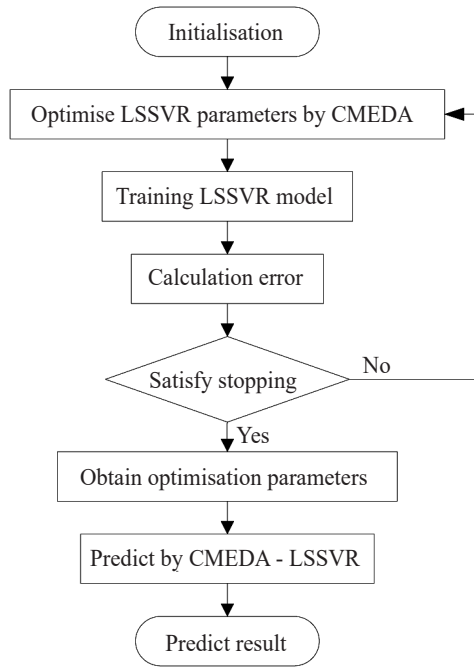


Fig. 2. Framework of CMEDA-LSSVR prediction procedure
CMEDA: chaotic mutation to improve the estimation of the distribution algorithm; LSSVR: least squares support vector regression

To compare the performance among CMEDA, particle swarm optimisation (PSO) algorithm, the genetic algorithm (GA) and grid search (GS) algorithm for optimising the parameters of the LSSVR model, the absolute percentage error (APE) values of these four models were calculated and the curves are shown in Fig. 5.

$$APE = \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (13)$$

where y_i is actual data value and $\hat{y}_i$ is prediction data value.

We also adopted the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) to further evaluate the prediction accuracy of each model.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (15)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (16)$$

where N is the number of the actual data set, y_i is the actual data value and $\hat{y}_i$ is the value of the prediction result.

Software

All algorithms in this paper were implemented in the MATLAB R2010a under Windows 7. The LSSVR

MATLAB toolbox codes were downloaded from <http://www.esat.kuleuven.be/sista/lssvmlab/>, free of charge. Others were custom developed which were the modification of the algorithm LSSVR.

Results and discussion

After numerous experiments, the original parameters of the CMEDA-LSSVR were given as follows: the max iteration number $G_{\max}=100$, population size $n=50$ and the chaotic control parameter $\lambda=3.845$, $np=25$, $k=2$. The fitness curve value is shown in Fig. 3, the curve was drawn by the best fitness of the training data and the best fitness was the minimal fitness of the individuals of each generation. It can be seen from the Fig. 3, that the fitness curve of the CMEDA optimisation algorithm is declining rapidly, converged after 45 iterations. PSO optimisation algorithm to adapt to the degree of decline in the curve is also faster, after about 40 iterations to converge, but its convergence value is greater than the CMEDA algorithm, that is, the prediction error is greater than the CMEDA algorithm. The convergence rates of GA and GS algorithms are relatively slower than the optimisation algorithms of CMEDA and PSO, and among which GS optimisation algorithm error is the biggest, followed by GA optimisation algorithm. This shows that the CMEDA algorithm is a good algorithm to find the two parameters of the LSSVR model.

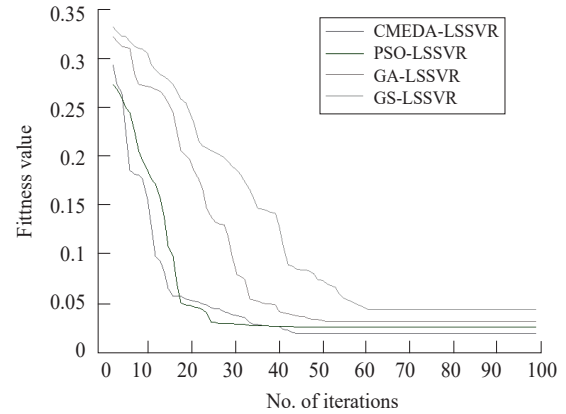


Fig. 3. Changing trend of the fitness function

The parameters of the LSSVR were optimised by the CMEDA algorithm and then we obtained: $\gamma=147.8315$ and $\sigma=1.9876$. To evaluate and compare the performance of the CMEDA-LSSVR, the PSO-LSSVR, GA-LSSVR and GS-LSSVR models were used for comparison. The PSO, GA and GS algorithm were used to optimise the two parameters of the LSSVR using the same training and testing data and we obtained the optimal parameters (137.263, 1.5412), (115.9864, 1.3672) and (126.295,

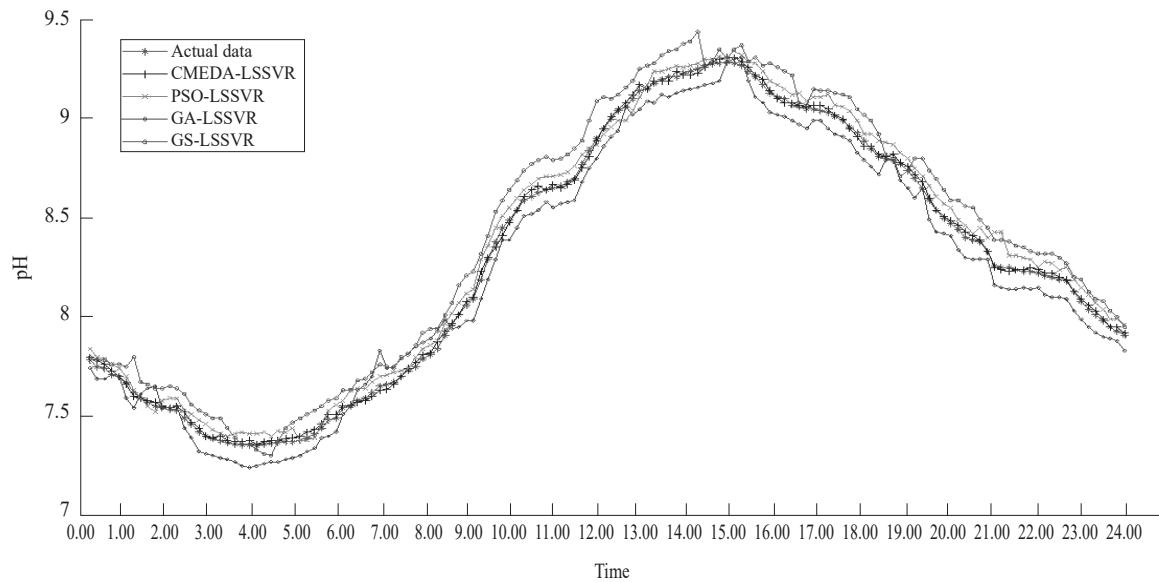


Fig. 4. Prediction results of the four models

CMEDA: Chaotic mutation to improve the estimation of the distribution algorithm; LSSVR: Least squares support vector regression; PSO: Particle swarm optimisation algorithm; GA: Genetic algorithm; GS: Grid search algorithm

1.7432), respectively. The prediction results of the four models and actual measured data are illustrated in Fig. 4.

As shown in Fig. 5, using the CMEDA-LSSVR model, the samples with the APE values less than 0.4% were 95% and the maximum APE was 0.48%. Using the PSO-LSSVR model, the samples with the APE larger than 0.4% were 81% and the APE values were mainly distributed in the range between 0.6% and 1%, with the maximum APE of 1.21%. With the GA-LSSVR model, the samples with the APE larger than 1% accounted for more than half and the maximum APE was 1.58%. With respect to GS-LSSVR model, the APE values of 80% samples were larger than 1% and the maximum APE was 2.43%. Obviously, the CMEDA-LSSVR model had the highest prediction accuracy among the four models.

One can easily observe from Table 1 that the RMSE, MAE and MAPE values of CMEDA-LSSVR model were much smaller than these parameters of PSO-LSSVR, GA_LSSVR and GS_LSSVR models. Compared with

Table 1. Precision statistic analysis results of four models

Model	RMSE	MAE	MAPE	T(s)
CMEDA_LSSVR	0.0381	0.0356	0.4059	9.8
PSO_LSSVR	0.0586	0.0539	0.6332	8.9
GA_LSSVR	0.0831	0.0777	0.9385	10.4
GS_LSSVR	0.1122	0.1038	1.2499	16.2

CMEDA: Estimation of the distribution algorithm improved by chaotic mutation; PSO: Particle swarm optimisation algorithm; GA: Genetic algorithm; GS: Grid search algorithm

the results of the other three models, the RMSE of CMEDA-LSSVR model was less by 34.9, 54.2 and 66%, respectively; the MAE of CMEDA-LSSVR model was less by 34, 54.1 and 65.7%, respectively and the MAE of CMEDA-LSSVR model was less by 35.9, 56.9 and 67.5%, respectively. The experimental results indicated that the prediction accuracy of CMEDA-LSSVR model was highest among the four models. The run time T(s) of CMEDA-LSSVR model was slightly longer than that of PSO_LSSVR model but shorter than those of the other

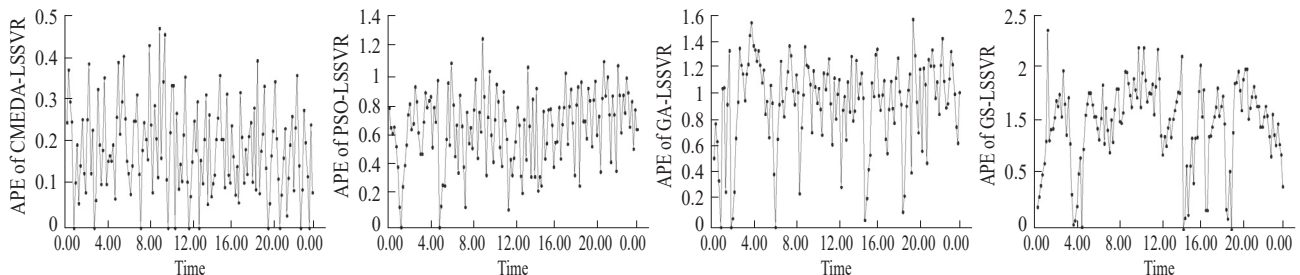


Fig. 5. Absolute percentage error (APE) values of four prediction models, calculated using the sampling data of 1 day

two models. This also indicates that it is feasible to find the optimal parameters with the use of the CMEDA algorithm. Moreover, the CMEDA-LSSVR is preferable to the other three predication models in performance.

In order to verify whether the different data sets have an effect on the prediction result of the CMEDA-LSSVR model, the measured data of 40 days were used as training and test data to test the capacity of the model. The experimental data of 40 days were divided into 5 groups: 7, 14, 21, 30 and 40 days, for data in each group with 80% of the data as training set and 20% of the data as test set. The absolute percentage errors of prediction results are: 0.63, 0.48, 0.34, 0.32 and 0.49%. It can be seen from the results that the use of nearly 3 weeks to a month's data as a data set can obtain a higher prediction accuracy. Because the pH value is influenced by dissolved oxygen, temperature, salinity, turbidity and light conditions and other factors, the use of fewer days data will make the pH forecast to produce relatively large fluctuations while the use of more days of data cannot reflect the recent weather conditions.

We also verified whether the CMEDA-LSSVR model was valid only for a particular set of data, or was general. The 30 days water quality data collected by the monitoring system was divided into 15 subsamples, that is, every two days of data as one subsample. Using 15-fold cross-validation, 1 subsample was used as the test data and the other 14 samples were used as training data and the cross validation was repeated 15 times and each subsample was verified once. The mean values of RMSE and MAE of prediction results of 4 models are shown in Fig. 6. As can be seen from Fig. 6, for a large sample, 4 prediction models using different training and test subsamples to predict pH value, the result of CMEDA-LSSVR model had the smallest error and GS-LSSVR had the biggest error. So it can be seen that

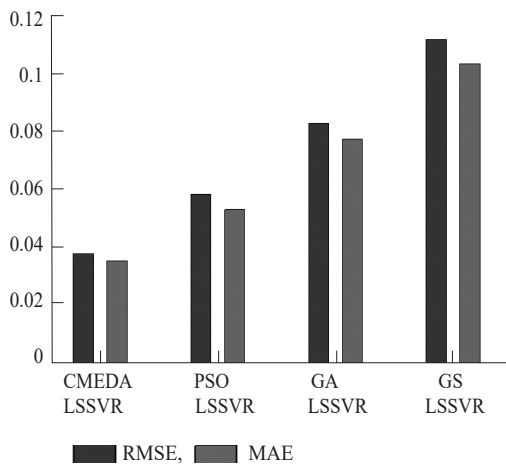


Fig. 6. Cross validation results of pH value prediction

CMEDA-LSSVR is suitable for the prediction of pH value and the classification of subsamples does not affect the prediction results. We downloaded the water quality data of the whole year in 2015 from the monitoring system database and repeated the above cross validation experiment with the remaining 11 months of water quality data. The average RMSE and MAE of prediction results were calculated and the results showed the same rule. Comparing with the other three models, the RMSE of CMEDA-LSSVR model were reduced by 32, 49 and 58%, respectively and MAE were reduced by 34, 51 and 59%, respectively, which indicated that the CMEDA-LSSVR model was suitable for crab pond pH value forecast.

To make it convenient for farmers to use the prediction model, we designed a graphical user interfaces (GUI) based on MATLAB which is shown in Fig. 7. Farmers could use a web browser to download the history data from the remote monitoring system for use as training data and testing data. To use this, the farmer has to open the GUI of the pH prediction model, press the “Load Data” button, load the training and testing data and press the “CMEDA_LSSVR” button, then the GUI will run the callback function and execute the CMEDA-LSSVR program. The prediction results of different time would be calculated and would be displayed on the GUI. Farmers could make a decision according to the prediction results, if the prediction results show that the pH value is lower than the lower limit of safety value/higher than the upper limit of safety value in the future, farmers could adopt appropriate control measures ahead of time. In our research, the safety interval of pH value was [7, 10] and from Fig. 7, it can be seen that the pH value of the pond was within safe levels up to next 10 h.

To improve the accuracy of the pH prediction model, we proposed the CMEDA-LSSVR model for pH value prediction in mitten crab culture. The RMSE, MAE and

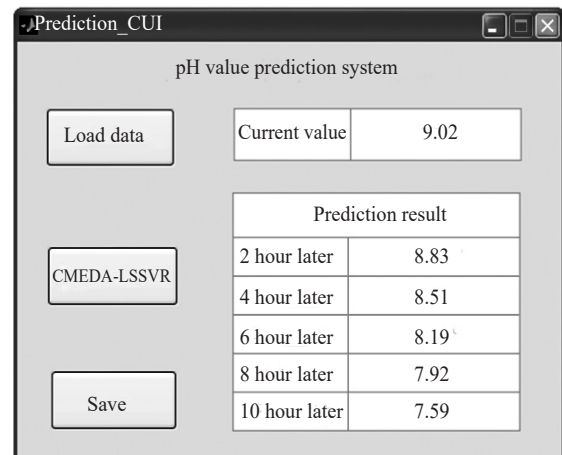


Fig. 7. GUI of pH prediction system

MAPE of the CMEDA-LSSVR model were 0.0181, 0.0156 and 0.4059%, respectively and overall results sufficiently demonstrated that the CMEDA-LSSVR model had a higher potential to predict the trend of pH than the other three models. We noted the sensitivity of the LSSVR parameters and in the future we plan to research other advanced optimal algorithms. Although many questions remain open, our research shows that the CMEDA-LSSVR model could be considered a suitable experimental model for future study on pH prediction.

Acknowledgements

The study was funded by the National Natural Science Foundation of China (31172243), the Agricultural Science and Technology Support Program of Jiangsu Province (BE2013402) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD, NO.6-2011).

References

- Alvarez Meza, A. M., Daza Santacoloma, G., Acosta Medina, C. D. and Castellanos Dominguez, G. 2012. Parameter selection in least squares-support vector machines regression oriented, using generalised cross-validation. *Dyna-Colombia*, 79(171): 23-30.
- Cai, C., Gu, X., Huang, H., Dai, X., Ye, Y. and Shi, C. 2012. Water quality, nutrient budget and pollutant loads in Chinese mitten crab (*Eriocheir sinensis*) farms around East Taihu Lake. *Chinese J. Oceanol. Limnol.*, 30(1): 29-36.
- Chen, W. J., Shao, Y. H., Deng, N. Y. and Feng, Z. L. 2014. Laplacian least squares twin support vector machine for semi-supervised classification. *Neurocomputing*, 145: 465-476.
- Cheng, Y. H., Wang, X. S. and Hao, M. L. 2010. An estimation of distribution algorithm with diversity preservation. *Acta Electron. Sin.*, 38(3): 591-597.
- Dong, J. G., Zhang, F. and Zhang, Y. H. 2014. A water quality changing prediction model for agricultural water-saving irrigation based on PSO-LSSVR. *Computer Modelling and New Technologies*, 18(7): 377-381.
- Ekici, L., Simsek, Z., Ozturk, I., Sagdic, O. and Yetim, H. 2014. Effects of temperature, time and pH on the stability of anthocyanin extracts: Prediction of total anthocyanin content using nonlinear models. *Food Anal. Meth.*, 7(6): 1328-1336.
- Gaya, M. S., Wahab, N. A., Sam, Y. M. and Samsuddin, S. I. 2014. ANFIS based effluent pH quality prediction model for an activated sludge process. *Conference ANFIS based effluent pH quality prediction model for an activated sludge process*, 2014, Johor Bahru, Malaysia, p. 538-542.
- Goodarzi, M., Freitas, M. P., Wu, C. H. and Duchowicz, P. R. 2010. pKa modeling and prediction of a series of pH indicators through genetic algorithm-least square support vector regression. *Chemometrics and Intelligent Laboratory Systems*, 101(2): 102-109.
- Hong, Y., Yang, X., Cheng, Y., Liang, P., Zhang, J., Li, M., Shen, C., Yang, Z. and Wang, C. 2013. Effects of pH, temperature and osmolarity on the morphology and survival rate of primary hemocyte cultures from the mitten crab, *Eriocheir sinensis*. *In Vitro Cell. Dev. Biol. Anim.*, 49 (9): 716-727.
- Igné, B., Reeves, J. B., McCarty, G., Hively, W. D., Lund, E. and Hurburgh, C. R. 2010. Evaluation of spectral pretreatments, partial least squares, least squares support vector machines and locally weighted regression for quantitative spectroscopic analysis of soils. *J. Near Infrared Spectrosc.*, 18(3): 167-176.
- Li, B. and Cai, L. 2013. Acidification prediction model of concrete based on grey system. *Kuei Suan Jen Hsueh Pao/J. Chin. Ceram Soc.*, 41(10): 1375-1380.
- Lin, K. P., Pai, P. F., Lu, Y. M. and Chang, P. T. 2013. Revenue forecasting using a least-squares support vector regression model in a fuzzy environment. *Information Sciences*, 220: 196-209.
- Liu, S.Y., Xu, L. Q., Jiang, Y., Li, D. L., Chen, Y. Y. and Li, Z. B. 2014. A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. *Engineering Applications of Artificial Intelligence*, 29: 114-124.
- Lou, W., Yang, K., Zhu, M., Wu, Y., Liu, X. and Jin, Y. 2014. Application of particle swarm optimisation-based least square support vector machine in quantitative analysis of extraction solution of yangxinshi tablet using near infrared spectroscopy. *J. Innov. Opt. Health Sci.*, 7(6): 323-524.
- Mahmoud, T. A. 2011. Adaptive control scheme based on the least squares support vector machine network. *Int. J. Appl. Math. Comput. Sci.*, 21(4): 685-696.
- Page, T. M., Worthington, S., Calosi, P. and Stillman, J. H. 2014. Effects of lowered pH on the exoskeleton mineralogy of porcelain crabs. *Integr. Comp. Biol.*, 54: E326-E326.
- Qiao, M. Y., Ma, X. P., Lan, J. Y. and Wang, Y. 2011. Time-series gas prediction model using LS-SVR within a Bayesian framework. *Mining Science and Technology*, 21(1): 153-157.
- Ritvo, G., Dassa, O. and Kochba, M. 2003. Salinity and pH effect on the colloidal properties of suspended particles in super intensive aquaculture systems. *Aquaculture*, 218(1-4): 379-386.
- Stephen, J. E., Kumar, S. S. and Jayakumar, J. 2014. Nonlinear modeling of a switched reluctance motor using LSSVM - ABC. *Acta Polytech. Hung.*, 11(6): 143-158.
- Sun, W. and Liang, Y. 2014. Research of least squares support vector regression based on differential evolution algorithm in short-term load forecasting model. *J. Renew. Sust. Ener.*, 6(5): 1-10.
- Tang, Z. H., Liu, J. P., Chen, Q., Gui, W.H. and Du, J. F. 2013. pH control in flotation process based on prediction model. *IET Control Theory and Applications*, 30(7): 885-890.

- Vapnik, V. N. 1995. *The nature of statistical learning theory*, Springer, New York, 188 pp.
- Wang, S. and Wang, Q. 2012. Prediction and dispatching of workshop material demand based on least squares support vector regression with genetic algorithm. *Information-an International Interdisciplinary Journal*, 15(1): 213-222.
- Xu, Z. and Xu, Y. 2016. A deterministic model for predicting hourly dissolved oxygen change: Development and application to a shallow eutrophic lake. *Water*, 8(2): 41
- Zheng, J., Shao, X., Gao, L., Jiang, P. and Qiu, H. 2014. A prior-knowledge input LSSVR metamodeling method with tuning based on cellular particle swarm optimisation for engineering design. *Expert Syst. Appl.*, 41(5): 2111-2125.