A water quality changing prediction model for agricultural water-saving irrigation based on PSO-LSSVR

Jian-Gang Dong*, Feng Zhang, Yong-Heng Zhang

School of Information Engineering, Yulin University, 719000, Yulin, China

Received 3 March 2014, www.tsi.lv

Abstract

In order to improve the prediction of early warning and agriculture information processing level of water quality for agricultural water-saving irrigation, using mathematics and information theory model to predict and estimate the possibility of future changes in water quality based on getting the quality data by using sensor device. The basic process, model for water quality prediction of agricultural water-saving irrigation, forecasting and early warning method of establishing process is designed. Finally, was using the PSO-LSSVR forecasting method to predict the water quality in the agricultural water-saving irrigation of water quality changes prediction. Simulation results show that the parameters of LSSVR were optimized by PSO algorithm, and overcome the cross validation to determine the influence of subjective factors of LSSVR parameters, has better prediction accuracy and generalization ability, its precision can satisfy the need for intensive irrigation production management.

Keywords: prediction model, PSO-LSSVR, water quality, water-saving irrigation, ZigBee

1 Introduction

It is one of the many field's applications of agricultural information processing method for agricultural forecast, It is using mathematics and information theory model to predict and estimate the possibility of future changes in water quality based on getting the quality data by using sensor device. This paper is discussed the agricultural water-saving irrigation forecast method, basic principles and basic steps, forecasting and early warning method is used, and was using the PSO-LSSVR algorithm prediction water quality of agricultural water-saving irrigation [1]. There were accurate prediction of the water quality parameters of agricultural water-saving irrigation according to the information and online monitoring data acquired, to grasp the variation of water quality in time and space and the development trend of the estimates and projections [2]. In advance of the irrigation water quality changes make proper identification, for agricultural water-saving irrigation management and water management departments in the water, making water quality planning adjustment measures and prevention of sudden deterioration of water quality events and provide scientific basis for decision making [3].

Prediction is refers to the various information and data on the history of investigation and statistics as the basis, from the point of view of things presented phenomena, using scientific methods and means, the possibility of future development of things projections and estimates, the development and changes of things in the future to make scientific analysis; thus the past and present to speculate about the future, by known to extrapolate, a science which reveals the trend of future development and law of objective facts or things. At present, there is extensive application in financial industry, commercial, meteorology, along with the development and popularization of agricultural technology, obtain more agricultural data, it's worthy to study agriculture forecast will become the future applications in the field [4].

It is in soil, environmental, meteorological data, growth, crop or animal conditions for agricultural production, fertilizer and pesticide, feed, aerial or satellite images and other practical agricultural data for agriculture forecast, based on economic theory, by means of mathematical model, projections and estimates the possibility of future development of the study object [5]. It is an important basis for scientific decision of precision fertilization, irrigation, sowing, weeding, pest control and other farming and agricultural production plan, supervise the implementation of the situation, but also improve the effective means of agricultural management. Therefore, agriculture prediction is one of the important technical means to support all aspects of agricultural production, sales activities based on the future agriculture Internet.

Agriculture forecast accuracy directly affects the quality of decision-making errors, and processing scheme thus, how to predict the various agricultural decision making process, and how to predict with high accuracy is an important problem in modern precision agriculture research. It's varied of prediction methods, all the effects of each prediction method cannot be completely contained to predict the target factor, in view of the complexity prediction target in agriculture and diversity,

* ***Corresponding author*** e-mail: donguncn@126.com

selection of prediction methods, it is the key what to establish to what mathematical model is more suitable for modern precision agriculture for prediction.

2 Water quality prediction method

In the agricultural water hydrology monitoring, the monitoring area wide, there are multiple items and unattended or duty personnel shortage, especially in the unexpected bad climate and environment is very difficult to ensure that the data transmitted to the monitoring centre in time, and may cause significant losses. For monitoring and early warning related hydrological data using wireless sensor networks, network mode is simple, time-saving, and real-time. Low cost, low power consumption, a sensor node dormancy and wake-up mechanism, can guarantee that the network can work stably for a long time in the field environment. Low cost, low power consumption, a sensor node dormancy and wake-up mechanism, can guarantee that the network can work stably for a long time in the field environment. The network comprises a water level gauge, gauge, anemometer and the gate level gauge according to the terminal node, can be installed in the rivers, reservoirs or farmland designated place of actual needs, work in the field of unattended [6]. ZigBee node through GPRS/CDMA wireless network or ADSL data collected will be sent to the hydrologic and water resources monitoring and management centre, has high reliability and expandability, combined with the superiority of the GPRS technology, is to achieve an ideal solution of Wireless Hydrological monitoring network.

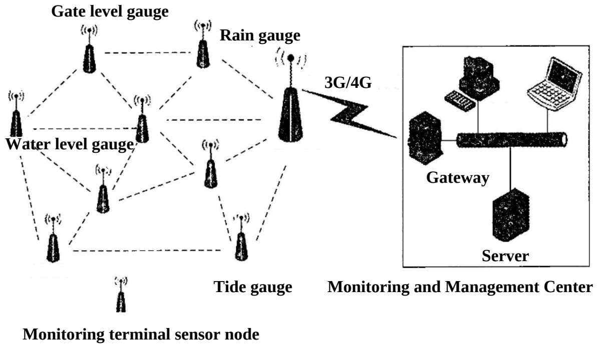


FIGURE 1 Agricultural water hydrology monitoring based on WSN

2.1 PARTICLE SWARM OPTIMIZATION ALGORITHM

Particle swarm optimization algorithm (PSO) is proposed by Kenney and Eberhart in 1995 population parallel search algorithm based on global optimization [7], through cooperation and competition between groups in the community to achieve optimal particle. Mathematical description of PSO: a population size is *n*, the *i* particles in m dimensional search space representation of $X_i = (X_{i1}, X_{i2}, ..., X_{ij}, ..., X_{im})$, flight speed is $V_i = (V_{i1}, V_{i2}, ..., V_{ij}, ..., V_{im})$, the optimal position of individual so far to search is $P_i = (p_{i1}, P_{i2}, ..., P_{ij}, ..., P_{im})$. The particle swarm optimal position is $P_{gbesti} = (p_{gbest1}, P_{gbest2}, ..., P_{gbestm})$. It can update the particle velocity and position according to the formula (1) and (2):

$$v_{ij}^{t+1} = \omega . v_{ij}^{t} + c_1 . r_1 . (p_{ij} - x_{ij}^{t}) + c_2 . r_2 . (p_{gbestj} - x_{ij}^{t}) , \quad (1)$$

$$x_{ij}^{t+1} = x_{ij}^{t} + v_{ij}^{t+1} , \quad (2)$$

where j=1,2,…,n; j=l,2,…,m; $c_1, c_2 > 0$ are respectively the individual learning factor and social learning factors; t is the current number of iterations, r_1 and r_2 are uniformly distributed random numbers in the range of [0,1]. ω is the inertia weight coefficient, used to control the effect of history on current speed. In order to balance the global and local search ability, make the ω along with the increase in the number of iterations decreases linearly, can significantly improve the performance of the PSO algorithm, it is given by

$$\omega = \omega_{\max} - t \times \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} . \quad (3)$$

In the formula (3), $\omega_{\max}$ is the initial inertia weight; $\omega_{\min}$ is the last inertia weight; $t_{\max}$ is the maximum number of iterations. Flight speed is $v_i \in [-V_{\max}, V_{\max}]$, the constraint conditions to prevent particle speed missed optimal solutions, through the improvement of the algorithm further improves the global searching ability of particle swarm.

2.2 LEAST SQUARES SUPPORT VECTOR REGRESSION ALGORITHM

Suykens proposed a least squares support vector regression (LSSVR) in 1999, is used to solve the problem of function estimation [8]. LSSVR is replaced by equality constraints and inequality constraints, the function of the error square and loss experience loss as training set, the traditional support vector machine in the solution of two quadratic programming problem is transformed into solving linear equation group, effectively improves the calculation speed and convergence precision, has better generalization performance. The mathematical model of the least squares support vector regression is given by:

$$\min J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^{1} \xi^T \xi , \quad (4)$$

$$s.t \quad y_i = \omega^T \varphi(x_i) + b + \xi_i . \quad (5)$$

In the formula, $X_i \in R^l$ and $Y_i \in R^l$ are the input and output vector system, $\xi_i \in R$ is the empirical error, b is offset, $C \in R^+$ is the regularization parameter, $\varphi(.)$ is a nonlinear mapping of the input space to the feature space. To solve the constrained optimization problems, Lagrange polynomial function of the dual problem is given by

$$L(\omega,b,\xi,a) = J(\omega,\xi) - \sum_{i=1}^{l} a_i(\omega^T, \phi(x_i) + b + \xi_i - y_i) . \tag{6}$$

In the formula (6), $a = [a_1, a_2, ..., a_l]^T$ is the Lagrange multiplier. According to the Karush-Kuhn-Tucher (KKT) conditions, respectively, for ω, b, ξ_i, a_j partial derivative, and let it equal to 0, linear system can be obtained as follows:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \to \omega = \sum_{i=1}^{l} a_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{l} a_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \to a_i = C\xi_i \\ \frac{\partial L}{\partial a_i} = 0 \to \omega^T \varphi(x_i) + b + \xi_i - y_i = 0 \end{cases} . \tag{7}$$

In the formula (7), to eliminate of ω, ξ_i, we can get the following linear equations

$$\begin{bmatrix} 0 & I^T \\ I & \Omega + C^{-1}E \end{bmatrix} * \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{8}$$

In the formula, $I = [1,1,...,1]^T$, E is a unit matrix of dimension $l \times l$; $y = [y_1, y_2, ..., y_l]^T$, $\Omega_{ij} = \phi(x_i)$, $\phi(x_j) = K(x_{i,} x_j)$ as the kernel function satisfying the Mercer conditions, the least square for the support vector regression model is given by

$$y = f(x,a) = \sum_{i=1}^{l} a_i k(x, x_i) + b . \tag{9}$$

3 Create the prediction model

3.1 THE LSSVR PARAMETER OPTIMIZATION BASED ON PSO

The study found the penalty factor C and kernel function parameter determines the performance of the LSSVR regression model. In order to improve the prediction performance of C and σ, it is the key of getting the best parameter combination. On the combination of parameter optimization of LSSVR model, there is no effective method, often through cross validation or gradient descent method, time-consuming and anthropogenic influence. Therefore, using the PSO algorithm for the parameters of LSSVR for automatic optimization, not only overcome the randomness of artificial selection, but also through particle fitness function settings, realize the automatic selection of objective parameters. Mean square error can directly reflect the performance of the LSSVR model (MSE) as the inverse function Fitness PSO algorithm's fitness (), its expression is shown as follows:

$$Fitness(C, \sigma^2, \varepsilon) = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y)^2}} . \tag{10}$$

In the formula (10), y_i, y were real and predicted values. Improved PSO LSSVR parameter optimization algorithm based on the procedure described as follows:

1) Particle swarm (C, σ) initialization. Set the number of particles size *n*, the maximum number of $t_{\max}$ iterations, range, the inertia weight ω particle velocity v limits, learning factor c_1, c_2 and other parameters, and a set of randomly generated initial particle velocity and position.

2) To train the LSSVR with the training set, by formula (10) are calculated for each particle's fitness value $Fitness(C, \sigma)$, then according to the particle's fitness value of individual extremum p_i and global extreme update p_{gbesti}.

3) By the formula (1), the formula (2) and formula (3), the speed and position of each particle is updated.

4) Checking algorithm termination conditions, such as the number of iterations is equal to T or optimal solution will not change, then the algorithm ends, output the optimal combination of parameters. Otherwise, proceed to step 3) optimization.

3.2 PREDICTION MODEL OF WATER QUALITY BASED ON PSO-LSSVR

The basic idea of the model for the nonlinear robust multi parameter water quality prediction system based on PSO-LSSVR is to make full use of PSO - LSSVR has the ability to establish the complex nonlinear relationship between the change of water quality and the reduced set of agricultural water-saving irrigation waters influence factors, mining internal variations of water quality, so as to realize the change of the historical factors make accurate prediction of future water quality change.

The model not only fully consider various factors with different times give different weights, optimized the parameters of LSSVR using improved PSO, effectively improve the accuracy of the prediction model and the generalization ability. The water quality prediction model building steps is shown in figure 2.

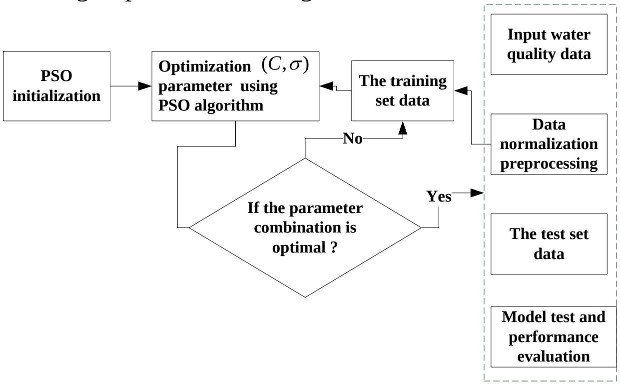


FIGURE 2 PSO-LSSVR water quality prediction model create step

4 The implementation of the model prediction

4.1 INTENSIVE WATER-SAVING IRRIGATION ECOLOGICAL ENVIRONMENT DATA SOURCE

In order to effectively forecast model for agricultural water saving irrigation dissolved oxygen inspection is presented in this paper, taking Yulin University network application demonstration base of agricultural water-saving irrigation reservoir ecological environment data as sample data, each sample including dissolved oxygen, temperature, pH, and rainfall index.

The sampling period is from 2014 May 10 to June 10th, sampled once every 30 minutes, a total of 561 samples, 500 samples collected before 7 days of training sets, the remaining 61 samples as a test set, taking the pond dissolved oxygen concentration on the next moment of quantitative prediction. The ecological environment monitoring part of the original data is shown in Table 1.

TABLE 1 Water saving irrigation ecological environment of original data

Time	Water temperature / °C	pH	Dissolved oxygen /(mg/L)	Rainfall / mm
06:00	25.132	7.08	6.6842	729
07:30	25.543	6.89	7.1231	456
08:10	26.113	7.23	6.9087	1295
09:20	26.453	7.12	7.9764	4324
12:00	30.211	6.98	7.9753	433
14:30	31.653	6.09	7.0984	543
17:30	32.756	6.34	6.0644	356
20:20	31.675	6.98	5.9745	453
22:10	30.087	6.91	4.5357	345
23:30	30.011	6.92	3.7543	344

4.2 DATA PRE-PROCESSING

Intensive water-saving irrigation dissolved oxygen is influenced by many factors, is a sequence of data changes with time, with different dimension.

If the direct use of the original data for combinatorial optimization for training the LSSVR parameters based on PSO algorithm, not only affects the learning speed, but also seriously restricts the accuracy and robustness of the predictive model, it is necessary to establish prediction model by type (10) of the original data pre-treatment, to reduce the data the dimension of different influence on the prediction model, so we can write

$$\bar{x}_m^n = \frac{x_m^n - \min(x_m^n |_{k=l}^l)}{\max(x_m^n |_{m=1}^l) - \min(x_m^n |_{m=1}^l)}, d = 1, 2, \ldots m . \qquad (11)$$

In the formula (11), l is the total sample, n is the dimension of the sample vector, $\bar{x}_m^n$ and x_m^n respectively the original data of water saving irrigation of the ecological environment and the normalized data.

4.3 ANALYSIS OF ALGORITHM AND PERFORMANCE

Algorithm using Matlab11 language programming, PSO algorithm is initialized: population size is $n = 50$, $c_1 = c_2 = 1.23$, the maximum number of iterations of $t_{\max} = 98$, the scope of the ω is set to [0.9, 0.35], particle velocity v limits [0.3, 8]. The dissolved oxygen, temperature, pH values as input of half an hour before, prediction of dissolved oxygen of the next moment values as output, the mean square error of prediction values of dissolved oxygen and the next time the actual value as the particle fitness function. According to the optimization algorithm is used to train the PSO-LSSVR LSSVR parameters based on PSO, the training times of t= 80, the best combination of parameters for LSSVR: σ =0. 0075, C=78.36. The combined prediction of dissolved oxygen concentration parameter generation PSO-LSSVR models, as is shown in figure 3.

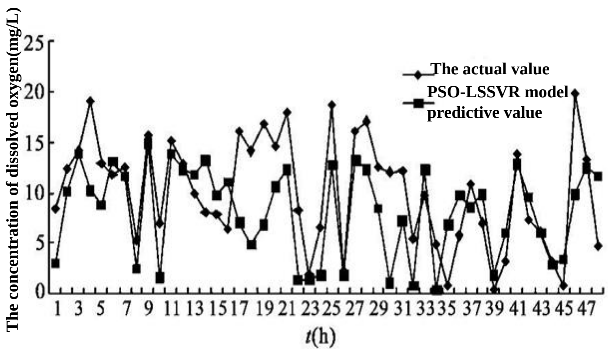


FIGURE 3 The comparison of prediction value and real value of dissolved oxygen based on PSO-LSSVR

It can be seen from Figure 3, the predicted curve are basically consistent with the measured curve, and with the combination of parameters of PSO algorithm for LSSVR optimization, overcomes the cross validation to determine the influence of subjective factors combination parameters of the LSSVR, has better prediction accuracy and generalization ability, the prediction accuracy can meet the need for intensive production of water-saving irrigation management.

5 Conclusions

This paper discusses the basic method for prediction of agricultural irrigation water quality, detailed analysis of the basic principles of prediction model of agricultural irrigation water quality prediction model, the selection principle, basic steps and methods of forecasting. The construction of the research method and the experimental method of model is predictive the PSO-LSSVR intensive agricultural water-saving irrigation based on water quality. The construction of intensive agricultural water-saving irrigation water quality prediction model of based on PSO-LSSVR is not necessarily the best forecasting model, the need for further research of intelligent computer technology and the improvement of the algorithm of support vector machine in the future research, in order to improve the robustness of water quality prediction.

Acknowledgments

This work is partially supported by the Funding Project for Department of Education of Shaanxi Province in 2013 #**2013JK1151**, Research and Cooperation Project for Department of Yulin city of (**#Gy13-16, #Sf13-08, #Sf13-23**). Thanks for the help.

References

[1] Avci E 2009Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm support vector machines: HGASVM *Expert Systems with Applications* **36**(2) 1391-402

[2] Babaoglu I, Findik O, Ulker E 2010 A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine *Expert Systems with Applications* **37**(4) 3177-83

[3] Cherkassky V, Ma Y 2004 Practical selection of SVM parameters and noise estimation for SVM regression *Neural Networks (S0893-6080)* **17**(1) 113-26

[4] Hsieh T J, Hsiao H F, Yeh W C 2012 Mining financial distress trend data using penalty guidedsupport vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm *Neurocomputing* **82** 196-206

[5] Üstün B, Melssen W J 2005 Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization *Analytical Chimica Acta (S0003-2670)* **544**(1-2) 292-305

[6] Yaoyu Feng, Zhimin He 2003 Optimization of agitation, aeration, and temperature conditions for maximum β-mannanase production *Enzyme and Microbial* **32**(2) 282-9

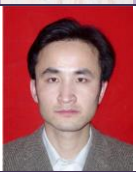
[7] Zhang Yong-Heng, Zhang Feng 2013 A New Time Synchronization Algorithm for Wireless Sensor Networks Based on Internet of Things *Sensors and Transducers* **151**(4) 95-100

[8] Cao X, Chen, Sun Y 2009 An Interface Designed for Networked Monitoring and Control in Wireless Sensor Networks *Computer Standards and Interfaces* **31**(3) 579-85

Authors

Dong Jian-Gang, born on November 28, 1974, in Shannxi Yulin

Current position, grades: associate professor in Yulin University.
University studies: MS degree in Computer science from Xidian University in 2009.
Scientific interest: Data mining technology, the Internet of Things applications.



Zhang Feng, born on June 26, 1980, in Shannxi Yulin

Current position, grades: associate professor in Yulin University.
University studies: MS degree in Computer science from Xidian University in 2009.
Scientific interest: Cloud integrated manufacturing technology, the modelling of complex systems, the Internet of Things applications.

Zhang Yong-Heng, born on October 25, 1968, in Shannxi Yulin

Current position, grades: associate professor in Yulin University.
University studies: MS degree in Computer science from Xidian University in 2010.
Scientific interest: Data mining technology, Mass data processing technology.