Neural, Parallel & Scientific Computations 15 (2007) 375-392

The Intelligent Control System of Flocculation Process Based on Genetic Wavelet Neural Networks for Sewage Treatment

Jingwen Tian[1,2] Meijuan Gao[1,2] Hao Zhou[2] and Jin Li[2]
[1]Department of Automatic Control, Beijing Union University
Beijing, China, 100101
[2]School of Information Science, Beijing University of Chemical Technology
Beijing, China, 100101

Abstract
Considering the issues that the flocculation process of sewage treatment is a complicated and nonlinear system, and it is very difficult to found the process model to describe it. The wavelet neural network has the advantages of both wavelet analysis and neural network, therefore it has the ability of strong nonlinear function approach and the ability of strong adaptive learning and it also has the feature of fast convergence and global optimization. Meanwhile the genetic algorithm has the global search ability. In this paper, an intelligent optimized control system based on genetic wavelet neural network is presented, the parameters of flocculation process are measured using multi sensors, then the control system can control the flocculation process real-time. The system is used in the sewage treatment plant. The experimental results prove that this system is feasible.

Keywords – genetic algorithm, wavelet neural network, sewage treatment, flocculation process, control system.

1. INTRODUCTION

Water is an important factor for the economic development and persistent development of society. With the development and the application of the sewage treatment technology, it is becoming the essential component parts that maintain the sustainable development of social economy. The instant development of the economy has brought out the more eminent environmental pollution problem. With the situation of water pollution is becoming seriously, the sewage treatment technology tends towards the track with the environment protection in the background. The crisis of water resource is now existing almost all over the world and this problem is especially severe in the cities with huge number of people. City sewage is the main reason for the water pollution of rivers and lakes. So it is a main reason which restricted the persistent development of many cities. At present, the process ratio of city sewage has become an important symbol to decide the civilization degree of a region.

By the recent 200 years, the process method of city sewage has changed from the original natural disposal and simple stair disposal into a profundity disposal method based on multiple advanced technologies, and then recycle. Sewage disposal and reuse is the effective measure for the exploitation of water resource. The objective of water reuse

Received February 15, 2007 1061-5369 $15.00 © Dynamic Publishers, Inc.

is to use the water disposed by membrane biology reactor to virescence, swash, complement for view water style etc, city sewage can be got in the nearby places, thus can avoid the water transport in distance, meanwhile, after disposal, the contamination is wipe off, not only the water is saved, but also the environment pollution is reduced. Sewage recycle has been executed in many areas lack water, and is considered to have obvious society, environment and economic benefit.

People pay more and more attention to the environment protection of city. For city with high population, the sewage treatment is more important. A great deal of methods for sewage monitor and treatment have been presented, and the sequencing batch reactor (SBR) is used universally at present. It combines the biology reactor and the secondary settling tank together, the sludge separation is finished by the gravity action in the secondary settling tank, and the separating efficiency depends on the settling capability of the sludge, the better the setting capability is, the higher the separating efficiency will be. It essentiality is to take the organism in the sewage as the culture medium, and sequentially culture the multiple kinds of animalcule colony in the condition of oxygenic, after the processes of agglomeration, adsorption, oxygenation decomposition, and deposition, the organism will be eliminated. SBR method has the properties of simple technics, high efficiency, good dephosphorization effect, strong defending capability of the sludge expands and high processing capability (Ren Min, Wang Wanliang, Li Tanwei, Guan Qiu, Yao Minghai, 2001).

The flocculation process not only performs an important role in sewage treatment but also decide the quality and the efficiency of sewage treatment. During the flocculation process, the flocculation agent was a medicament that is added in order to wipe off the suspended substance and colloid. The flocculation agent can make the feculence change into flocky precipitate and accelerate the process of purify, the alkali can adjust the PH value of water and the chlorine can sterilize. The dosage of flocculation agent is related to many factors such as degree of turbidity, temperature of water, environment changing and so on.

Above the flocculation process some research works have been done (Wu Daoji et al., 2000; Shen H, Iha H et al., 2006; Kaseamchochoung, Chudapak et al., 2006; Kobayashi O, Suda H, Ohtani T, Sone H, 1996). All the results show that the whole flocculation process has the feature of multivariate, nonlinearity, time variant nature and randomness, it is a complicated and nonlinear system, to which the control process is quite complicated, and traditional control method could not satisfy the real-time control on the flocculation process of sewage disposal.

The flocculation control technology is the key tache for the water purifying, an inadequate control will not only make the determined water quality can not be reached, but also result in the waste of medicament (XIANG Zhuang-li, 2004; Quan Jiping, Huang

Xiaodong, Xiao Weigui, et al., 2005). At present, it is very difficult to found the process model to describe the flocculation process. In generally the dosage of the flocculation agent are decided by the experienced workers. The shortness of this method is that it cannot satisfy the need of continuous running, so we medicament quantity to be added cannot be adjusted in time, and the supplied medicament ratio to be added is only the referent value for the real added quantity, for it not only results in inaccuracy but also the lag of detected value. The intelligent control is mainly used for solving the complex control problems that are hard to be solved by traditional method. Fuzzy control, NN control, expert control are the main branches of intelligent control. And the NN control is especially proper to used for the control of multi-variables with high uncertainty and high non-linear property (Special Issue on Fuzzy Logic and Neural Networks, 1992).

The error back propagation BP network is a new technique in recent years, its ability to approach nonlinear function has been proved in theory also have been validated in actual applications (Jiao Licheng, 1995). BP network is one kind of global approach network and the weight value of network is adjusted to every group data of outputs and inputs while the network is trained, this conduces the speed of network learning slowly. On the other hand, the BP has some problems such as converge to local minimum, the overfitting and the structure of network is always decided by experience because it doesn't have a good guiding theory. So the model built by BP may create mistake in the optimized control.

Wavelet neural network is new kinds of network based on the wavelet transform theory and the artificial neural network (Pati Y C, Krishnaprasad P S, 1993). It full utilizes the good localize character of the wavelet transformation and combines the self-learning function of the neural network. Its characters are: the confirmation of the wavelet base and the whole network has dependable theory warranty, so it can avoid the sightlessness in structural design of BP network; The learning object function which concerned to weights is convex, therefore the global minimum solution is exclusive; It has the ability of strong adaptive learning and function approach (Zhang Q, Benveniste A, 1992). Genetic algorithm is also a new optimum algorithm developed fast recently, it has the global search ability (Zhao Zhenning, Xu Yongmao, 1996; Grefenstetle J, 1986). The intelligent control strategy doesn't need an intact mathematical model of object. It can be used to control the complicated system that has the nonlinearity and the time variant nature. This provides a good method of resolution for the flocculation process control in sewage treatment. The genetic wavelet neural network has the ability of strong function approach and fault tolerance and the ability of strong adaptive learning. Meanwhile it can overcome the disadvantages of the BP algorithm. The genetic wavelet neural network has the simple implementation process and fast convergence rate, it adapts to intelligent control. So the genetic wavelet neural network was applied to the intelligent control of flocculation process in this paper.

2. PRINCIPLE

2.1. BP network

The neurons are arranged as some layers in BP network, the network composed by one input layer and one or more hidden layers and one output layer. The learning course of network includes two courses, one is the input information transmitting forward directed and another is the error transmitting backward directed. In the transmitting forward direction, the input information goes to the hidden layers from input layer and goes to the output layer. If the output of output layer is different with the wishful output result then the output error will be calculated, the error will be transmitted backward directed then the weights between the neurons of every layers will be modified in order to make the error become minimum.

A three layers BP network is shown as follow Figure 1. The numbering of input layer is i, the numbering of hidden layer is j, the numbering of output layer is k.

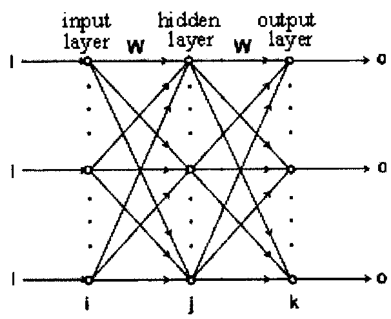

Figure 1. Structure of BP neural network

Then the input of the j neuron of hidden layer is:

$$net_j = \sum_i w_{ji} o_i \tag{1}$$

The output of the j neuron is:

$$o_j = g(net_j) \tag{2}$$

The input of the k neuron of output layer is:

$$net_k = \sum_j w_{kj} o_j \tag{3}$$

Corresponding output is:

$$o_k = g(net_k) \tag{4}$$

Where, g is Sigmoid function,

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

The key of BP network is the error transmitting backward directed of learning course. The course is accomplished through minimize an object function that is the error sum of squares between the actual output of network and the expectant output. Using the gradient descent algorithm we derive the computing formula.

In the learning course, supposed the expectant output of the k neuron of output layer is t_{pk}, the corresponding actual output of network is o_{pk}, then the average error of system is

$$E = \frac{1}{2p}\sum_{p}\sum_{k}(t_{pk} - o_{pk})^2 \tag{5}$$

Where, p is the training samples number.

In order to the express convenience, omit the subscript p, the formula (5) becomes as follow:

$$E = \frac{1}{2}\sum_{k}(t_k - o_k)^2 \tag{6}$$

Where, E is the object function.

According to the gradient descent algorithm, we derive the adjustment value of every weighting as follow:

$$\Delta w_{kj} = \eta(t_k - o_k)o_k(1 - o_k)o_j \tag{7}$$

$$\Delta w_{ji} = \eta\delta_j o_i \tag{8}$$

Where, η is the rate of learning;

$$\delta_j = o_j(1 - o_j)\sum_{k}\delta_k w_{kj}$$

$$\delta_k = (t_k - o_k)o_k(1 - o_k)$$

The bigger the rate of learning η became, the more the adjustment value of every weighting, this can accelerate the training course of network, but the result can generate oscillation. In order to avoiding the oscillation when increase the rate of learning η, adding a momentum term in the formula (7) and (8), namely:

$$\Delta w_{ji}(n+1) = \eta\delta_j o_i + \alpha\Delta w_{ji}(n) \tag{9}$$

Where, α is the proportionality constant.

Through the BP network training, satisfy the accuracy requirement, then the interconnect weighing between every nodes are ascertained. Here, the trained network can identify and predict the unknown sample.

2.2. Wavelet neural network

The wavelet neural network (WN) is a neural network model which based on wavelet analysis. It replaces the common non-linear sigmoid function with non-linear wavelet basic function (Zhang Q, Benveniste A, 1992), the corresponding weights from the input layer to the hidden layer and the threshold value of the hidden layer are replaced respectively by the scale parameter and the translation parameter of the wavelet. The output of network is the linear superposition of the chosen wavelet base, namely the output of output layer is a linear neuron output.

To $\forall \psi \in L^2(R)$, if $\hat{\psi}(\omega)$ which is Fourier transform of $\psi(t)$ satisfy condition:

$$c_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty \tag{10}$$

Then we name the $\psi(t)$ is a basic wavelet or wavelet generating function. One function family can be build by telescoping and translation the $\psi(t)$ as follow:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi(\frac{t-b}{a}) \quad a,b \in R, \quad a \neq 0 \tag{11}$$

Here, $\psi_{a,b}(t)$ is named as wavelet, a is the scaling parameter, b is the translation parameter.

Wavelet transform of signal $f(t) \in L^2(R)$ is defined as follow:

$$W_f(a,b) = < f, \psi_{a,b} > = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \psi(\frac{t-b}{a}) dt \tag{12}$$

According to the wavelet transform principle: In the Hilbert space, select a wavelet generating function $\psi(x_{t-1}, x_{t-2}, \cdots, x_{t-p})$, make it satisfy the admissibility condition:

$$c_\psi = \int_{L^2} \frac{|\hat{\psi}(\omega_{x_{t-1}}, \omega_{x_{t-2}}, \ldots \omega_{x_{t-p}})|}{|\omega_{x_{t-1}}|^2 + |\omega_{x_{t-2}}|^2 + \cdots + |\omega_{x_{t-p}}|^2} d\omega_{x_{t-1}} \ldots d\omega_{x_{t-p}} < +\infty \tag{13}$$

Here, $\hat{\psi}(\omega_{x_{t-1}}, \omega_{x_{t-2}}, \ldots \omega_{x_{t-p}})$ is the Fourier transform of $\psi(x_{t-1}, x_{t-2}, \ldots x_{t-p})$.

Doing the telescoping, translation and rotational transform to the $\psi(x_{t-1}, x_{t-2}, \ldots x_{t-p})$, we can obtain the wavelet basic function as follow:

$$\psi_{a,\theta,\bar{b}}(x_{t-1}, \ldots, x_{t-p}) = a^{-1} \psi(a^{-1} r_{-\theta}(x_{t-1} - b_{x_{t-1}}, \ldots, x_{t-p} - b_{x_{t-p}})) \tag{14}$$

Be abbreviated to $\psi_{a,\theta,\bar{b}}(\cdot)$. Here $a \in R$, and $a \neq 0$ is the scaling parameter, $\bar{b} = (b_{x_{t-1}}, \ldots, b_{x_{t-p}}) \in R^p$ is the translation parameter.

The rotating vector is defined as follow:

$$r_{-\theta}(x_{t-1}, \cdots x_{t-i}, \cdots x_{t-j}, \cdots, x_{t-p}) = x_{t-i} \cos\theta - x_{t-j} \sin\theta \quad 1 \leq i \leq j \leq p \tag{15}$$

The $\psi_{a,\theta,\bar{b}}(\cdot)$ can satisfy the framework property of the function space by proper selecting the scaling parameter, translation parameter $a > 0, \bar{b}, \theta$.

$$A \| f \|^2 \leq \sum_{a,b} |< \psi_{a,\theta,\bar{b}}, f >|^2 \leq B \| f \|^2 \quad 0 \leq A \leq B < \infty \tag{16}$$

Rearrange the set of wavelet functions $\{ \psi_{a,\theta,\bar{b}}(\cdot) \}$ into $\psi_1(\cdot), \psi_2(\cdot), \ldots, \psi_n(\cdot)$, and replace the common non-linear sigmoid function of the single hidden layer with it, we can obtain the corresponding wavelet neural network. The topology structure diagram of the wavelet neural network is shown as Figure 2.

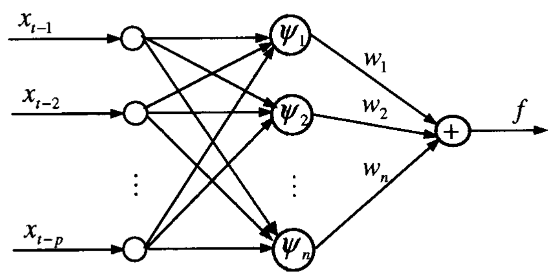

Figure 2. The structure diagram of WN

2.3. Genetic algorithm

The genetic algorithm (GA) is a kind of self-adapting heuristic global search algorithm which derived from imitating the thought of natural biological evolution. In nature, it is a cycle process made up of reproduction-crossover -mutation operators. In the process of searching for the global optimum solution, GA needs neither the information of gradient nor the calculus computing, it can find out the global optimum solution or near-optimal solution in the solution space with high probability only by operating the reproduction-crossover -mutation operators. Thereby, it could reduce the probability of getting into the local minimum efficiently.

The reproduction operator reproduces the individuals to the new colony according to the probability in proportion as their adaptive value. After reproduction, the preponderant individuals are preserved and the inferior individuals are weed out, and the average fitness degree of the colony is increased, but the variety of colony is loss at the same time. The action of reproducing operator is to realize the principle of winner priority for preserving predominance and natural selection, and make the colony converge on the optimum solution. The crossover operator first selects two individuals stochastically according to the certain exchanging probability P_c , then it can produce two new individuals by exchanging parts of chromogene stochastically. The genetic algorithm can generate filial generation colony which have higher average fitness and better individuals through the reproduction and crossover operators, and make the evolutionary process proceed to the optimum solution. The mutation operator changes several bits of the chromosome string stochastically with a small probability P_m, namely turn 0 to 1 and 1 to

0. The mutation operator is very important to recoup the loss of colony diversity and it can avoid the algorithm getting into the local minimum.

AGA is a kind of GA that has scale reproduction and self-adaptive crossover and mutation operations. In the process of searching for the optimum parameter, AGA changes the crossover probability and mutation probability adaptively according to the different condition of individuals in order to keep the diversity of colony and prevent the premature convergence, further it can enhance the calculating speed and precision of the algorithm (Huang Xiuxuan, Zhu Xuefeng. 1998).

$$P_c = \begin{cases} k_1 (f_{\max} - f')/(f_{\max} - f_{avg}) & \text{if } f' > f_{avg} \\ k_3 & \text{if } f' < f_{avg} \end{cases} \tag{17}$$

$$P_m = \begin{cases} k_2 (f_{\max} - f)/(f_{\max} - f_{avg}) & \text{if } f > f_{avg} \\ k_4 & \text{if } f < f_{avg} \end{cases} \tag{18}$$

Here, $f_{\max}$ is the biggest fitness of colony, f_{avg} is the average fitness of colony, f' is the bigger fitness of two strings used for exchange, f is the fitness of the individual to mutate.

Generally: $k_1 = k_3 = 1$, $k_2 = k_4 = 0.5$. At practical application, the value of P_c is often in range 0.5-1.0, and the P_m in range (0.005-0.05).

3. SYSTEM ARCHITECTURE AND WORKING PROCESS

3.1. System architecture

Confessedly, the aim of a control system is to obtain the anticipant output by ascertaining proper input of controlling factors. The intelligent control system based on the genetic wavelet neural network presented in this paper is to use the genetic wavelet neural network as the controller. We can build the model by training the network and make the peculiarity of network model same as the inverse peculiarity of controlled object, thereby control the controlled object (Li Shiyong. 1996). Following Figure 3 shows the configuration of the intelligent control system.

The control system contains two genetic wavelet neural networks WN1 and WN2 with the same structure, where the WN1 is feedforward controller and the WN2 is the identification model of the controlled object. The WN1 requires the anticipant output y_d of control system as the input of wavelet neural network WN1 and the output of WN1 is the input u of the controlled object. The output response of the control system is y in drive of u; The input of WN2 is the input u of the controlled object and the output of WN2 is the output response y of the controlled object. In this system, first WN2 is

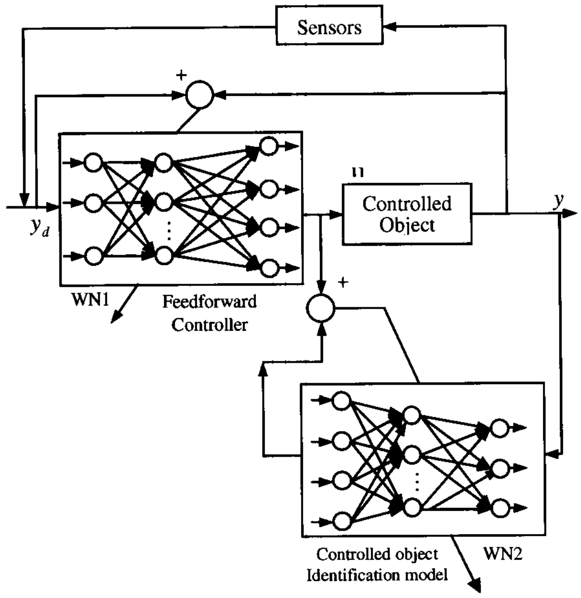

Figure 3. The configuration of the intelligent control system

trained off-line using all known samples and then get the in-out property of the controlled object. The network weights of the trained WN2 can be taken as the initial network weights of the WN1. Then the WN1 is trained using the back-propagation error of WN2, namely using the error between the anticipant output y_d and the real output y of control system to adjust the every network weights of WN1 to make the real output of system approach the anticipant output y_d. The WN1 can be trained on-line and revised while the control system is running.

3.2. The wavelet neural network use for intelligent control of the flocculation process

The WN1 and WN2 used in the control system are all the wavelet neural network with the same number of neurons in input layer, hidden layer and output layer. The wavelet neural network structure use for intelligent control of the flocculation process is shown as Figure 4.

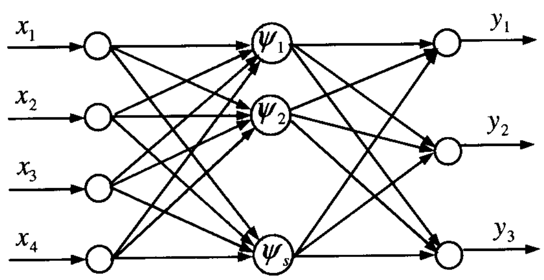

Figure 4. The wavelet neural network use for intelligent control of the flocculation process

The input layer is composed of four neurons that they express the dosage of the flocculation agent, alkali, chlorine and the temperature of water respectively and use the x_1, x_2, x_3 and x_4 express. The output layer is composed of three neurons that they express the degree of turbidity, PH value of the water and the flocky precipitate and use the y_1, y_2 and y_3 express. If we use the $I_i^{(j)}$ and $O_i^{(j)}$ respectively express the input and output of the i neuron of the j layer, then the in-out relationship of every layer of wavelet neural network can be expressed as follow:

The first layer of network puts the input into the network:

$$O_i^{(1)} = I_i^{(1)} = x_i \quad (i = 1,2,3,4) \tag{19}$$

The second layer (hidden layer) composed of wavelet basic function. Supposed the number of the wavelet basic function is s, then the in-out relationship of the hidden layer is expressed as follow:

$$\begin{cases} I_i^{(2)} = [O_1^{(1)}, O_2^{(1)}, O_3^{(1)}, O_4^{(1)}]^T \\ O_i^{(2)} = \psi_i(I_i^{(2)}) = \psi_i(\dfrac{I_i^{(2)} - b}{a}) \end{cases} \quad (i = 1,2,\cdots,s) \tag{20}$$

The network output obtained by the third layer (output layer):

$$y_i = O_i^{(3)} = I_i^{(3)} = \sum_{j=1}^{s} w_{ij} O_j^{(2)} \qquad (i = 1,2,3) \tag{21}$$

3.2.1 Ascertain of the number of wavelet basic function

The in-out relationship of every layer of wavelet neural network could be obtained from the formula (19), (20) and (21). The number of the wavelet basic function is a key parameter, it can influence the capability and the operational speed of the network. Moreover, through research, we know that the number of the wavelet basic function will increase acutely in company with the dimension increase. So it is the key of the flocculation process optimizing control to decrease the number of the wavelet basic function. We adopt a method of reduce the number of the wavelet basic function by analysis the sparsity property of sample data in this paper.

The sample data that we can obtain is bounded and sparse in most high-dimension actual problem. Maybe that some wavelet basic function do not cover any sample data in the wavelet network. These wavelet basic functions have no use to reconstructed function, so they could be deleted. Therefore, the first step of the algorithm is to delete the wavelet basic function which does not cover any sample data.

Suppose that the wavelet function of the network is compactly supported or approximately compactly supported, the wavelet basic function which does not cover any sample data would be deleted in its supported set.

If the function $\psi_i(x)$ is compactly supported, then its supported set S_i can express as follow:

$$S_i = \{x \in R^d : \psi_i(x) \neq 0\} \tag{22}$$

Here, d is the dimension of input space.

If the function $\psi_i(x)$ is not compactly supported, but it can tend to zero rapidly, the function $\psi_i(x)$ is referred to as the approximately compactly supported. Then its supported set S_i can express as follow:

$$S_i = \{x \in R^d : |\psi_i(x)| > \varepsilon \max_x |\psi_i(x)|\} \tag{23}$$

Here, ε is a tiddly positive number that preliminary definition.

Using (X,Y) express the sample data set which include M pair of sample data. For every $x_k \in X$, the order number assemblage I_k of the wavelet basic function which include the x_k in the compactly supported can express as follow:

$$I_k = \{i : x_k \in S_i\} \quad (k = 1,2,\cdots,M) \tag{24}$$

Then the number of the wavelet basic function is as follow:

$$T = \{\psi_i : i \in I_1 \cup I_2 \cup \cdots \cup I_M\} \tag{25}$$

Suppose that L is the number of the wavelet basic function of T, then

$$T = \{\psi_1, \psi_2, \cdots, \psi_L\} \tag{26}$$

Actually, some items of T have no use to reconstruct the original function. Because we only take the input sample data X into account, but do not consider the output sample data Y. So the second step of algorithm is to delete the useless items in T.

Deleting the useless items is equal to choose a subset in T, just make its span space as close as possible to the output vector Y. The matter could be divided into two subproblem: one is how to ascertain the size of the subset, the other is how to choose the item in the subset.

For the problem of choosing the item in the subset, we suppose the subnet size is s, finally we choose s items from the T. In order to attain this object, firstly, we choose an item of most eligible sample data from the T, then choose the fittest item from the residual terms repeatedly. For calculation convenience, the item that is chosen later should be perpendicular to the former ones.

For the problem to ascertain the size of the subset s, as the size of the T is L, so the value of s should be a number between 1 and L. We test every feasible value, namely choose s wavelet basic function to construct the wavelet neural network with the algorithm mentioned above for every feasible s value. Then evaluate the performance of the network using the least mean square deviation method. The s value of the network which has the least value of mean square error (MSE) between the network output and the sample actual output corresponding is the s value that we need gain.

The mean square error of the wavelet neural network is defined as follow:

$$E = \frac{1}{2M} \sum_{i=1}^{M} \sum_{k=1}^{K} (f_k(x_i) - y_{ik})^2 \tag{27}$$

Here, M is the amount of training samples;

K is the number of the neuron of network output layer;

$f_k(x_i)$ is the wavelet network output of output layer k neuron corresponding to the i sample.

y_{ik} is the actual output of the output layer k neuron corresponding to the i sample.

The number of wavelet basic function that we expected is as follow:

$$s = \arg\min_{s=1,2,\cdots,L} MSE \tag{28}$$

3.2.2 Learning algorithm

The generic algorithm with adaptive and floating-point code is combined with the gradient falling algorithm to train the wavelet network. The generic algorithm with adaptive and floating-point code denote the parameters directly with decimal-coding instead of binary-coding, thus, (1) it can avoid the encode difficulty caused by the ambiguity of the numeric area of the network. (2) Cancelled the process of encode and decode, so enhanced the learning speed of algorithm. (3) The importing of decimal numeric string can enhance the computational accuracy greatly under the circumstance of the length of the numeric string is invariable.

When the number of wavelet basic function was ascertained, the output of wavelet neural network was expressed as follow:

$$\hat{f}(\cdot) = \sum_{i=1}^{n} w_i \psi_{a,\theta,\bar{b}}(x_{t-1}, x_{t-2}, \ldots, x_{t-p}) = \sum_{i=1}^{n} w_i \psi_i(\cdot) \tag{29}$$

$$\text{or } \hat{f}(\cdot) = \sum_{i=1}^{n} w_i \psi_i(\cdot) \tag{30}$$

Here, w_i is the weight between the hidden layer node and the output layer node; $\psi_{a,\theta,\bar{b}}(\cdot)$ is the output value of the hidden layer node.

Had authenticated (Xie Meiping. 1998), the error limit of approaching the nonlinear function $f(\cdot)$ using the formula (29) or (30) was $O(n^{-\frac{1}{2}})$.

Gather together all the parameters of formula (29), and go by the general name of ϕ, replaced the $\hat{f}(\cdot)$ of formula (29) use $f_\phi(\cdot)$. We adopted the learning algorithm based on the gradient descent and generic algorithm with adaptive and floating-point code to train the wavelet neural network.

For the gradient descent algorithm, the object function was constructed as follow:

$$E = c(\phi) = \frac{1}{2}\sum_{p}[f_{\phi}(\cdot) - y]^2 \tag{31}$$

Here, p is the number of the training samples.

Recurrence decrease the formula (31) with the in-out data pair. Through the gradient algorithm of every measured value, the parameter ϕ was attenuated along the gradient direction of function.

$$c_K(\phi) = \frac{1}{2}[f_{\phi}(\cdot) - y_k]^2 \tag{32}$$

Our goal was to ascertain $w_i, a_i, b_i, \gamma_{-\theta}$, and make the fitting of that between the predicted value $f_{\phi}(\cdot)$ sequence of the wavelet network and the actual value y_k sequence was optimum. Here, $w_i, a_i, b_i, \gamma_{-\theta}$ can be optimized by the least square error energy function formula (32).

We adopted the Morlet wavelet as the excitation function of the hidden layer nodes of wavelet neural network, $\psi(t) = (e^{-i\omega_0 t} - e^{-\omega_0^2/2})e^{-t^2/2}$. When $\omega_0 \geq 5$, $e^{-i\omega_0/2} \approx 0$, the second item can be ignored, generally the approximate representation was $\psi(t) = e^{-i\omega_0 t}e^{-t^2/2}$.

Let $\psi(x) = d\psi(x)/dx$, $e_K = f_{\phi}(\cdot) - y_K$, $Z_i = a(x - t_i)$, then the partial differential of the function formula (32) corresponding to every component of every parameter vector ϕ was as follow:

$$\frac{\partial c}{\partial f} = e_K \tag{33}$$

$$\frac{\partial c}{\partial w_i} = e_K.\psi(Z_i) \tag{34}$$

$$\frac{\partial c}{\partial b_i} = -e_K.w_i.\gamma_{-\theta}.\psi'(\gamma_{-\theta}(Z_i)) \tag{35}$$

$$\frac{\partial c}{\partial a_i} = -e_K.w_i.\gamma_{-\theta}^2.\psi'(\gamma_{-\theta}(Z_i)) \tag{36}$$

$$\frac{\partial c}{\partial \gamma_{-\theta}} = -e_K.w_i.(x - t_i).\psi'(\gamma_{-\theta}(Z_i)) \tag{37}$$

Here, the modified value of weight is: $\Delta w_i = \frac{\partial c}{\partial w_i}$

The computation procedure was described as follow:

1) Initialization w_i, b_i, a_i and $\gamma_{-\theta}.i = 0$;

2) Compute the parameter in the step 1) according to the formula that from (34) to (37);

3) Substitute these parameters that calculated in the step 2) into formula (29), and compute $\hat{f}(\cdot)$;

4) Compute the error $err = \sqrt{\sum_{i=1}^{n}(f_i(\cdot) - y_i)^2 / \sum_{i=1}^{n} y_i}$;

5) If the error could satisfy the accuracy requirement and stop; else $i = i+1$, turn to the step 2).

For the generic algorithm with adaptive and floating-point code, the GA regards each weights of the network w_i, b_i, a_i and $\gamma_{-\theta}$ as a chromosome, and the aggregation of all the weights w_i, b_i, a_i and $\gamma_{-\theta}$ as an individual, and a large number of individuals will be generated in the initialization phase, which is called colony. The adaptation function of GA is constructed as follow:

$$f = \begin{cases} C_{\max} - E & E < C_{\max} \\ 0 & E \geq C_{\max} \end{cases} \tag{38}$$

Here, the $C_{\max}$ can be the maximum value E of evolutionary process.

In order to ensure the stability and global convergence of the algorithm, we adopt the best reserve mechanism in the selecting operation of GA, firstly according to the roulette selecting mechanism to select, then the most fitness individual of current solution is reproduced to the next generation colony, in order to ensure the final result that obtained as soon as the GA ends is the most fitness individual of every generation appear.

The steps of algorithm which combining the generic algorithm with adaptive and floating-point code with the gradient falling algorithm to train the wavelet network can be described as follow:

Step1: Random generate N groups initial network weights and parameters w_i, b_i, a_i and $\gamma_{-\theta}$ from different space interval of real number, and regard them as the initial colony;

Step2: Preliminary train these N groups initial weights and parameters w_i, b_i, a_i and $\gamma_{-\theta}$ separately using gradient falling algorithm, if there are at least one group satisfied the accuracy requirement after training, then the algorithm end, else turn step3;

Step3: Define the numeric area respectively according to the upper limit and inferior limit of the N (N is odd number) groups network weight and parameters w_i, b_i, a_i and $\gamma_{-\theta}$ which have been preliminary trained, random generate $r \times N$ groups new network weights and parameters in the numeric area, these new weights and parameters w_i, b_i, a_i and $\gamma_{-\theta}$ in conjunction with the N groups trained weights and parameters compose an holonomic colony, there are $(r+1) \times N$ groups network weight and parameters;

Step4: Execute reproduction, crossover and mutation adaptive genetic operation on the $(r+1) \times N$ groups weights and parameters;

Step5: If there are at least one group weights and parameters can satisfy the accuracy requirement after step4, the algorithm end, else select N groups better weights and parameters from the $(r+1)\times N$ which have been exerted on the generic algorithm with adaptive and floating-point code, and turn step2.

The algorithm flow is showed as Figure 5.

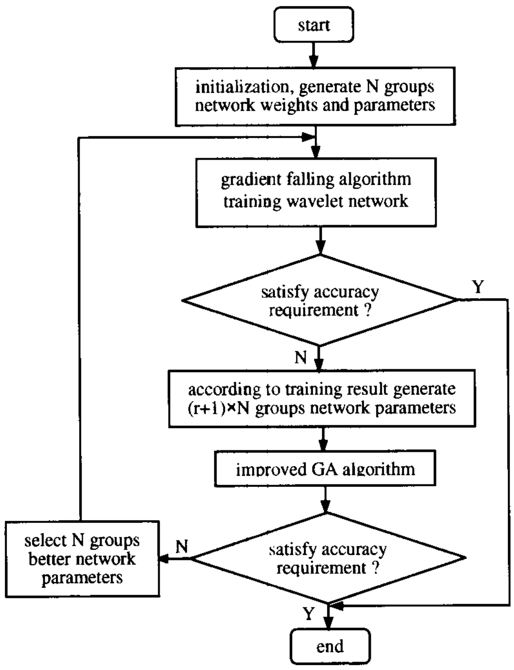

Figure 5. Flow chat of genetic wavelet network

3.3. Working process

In the working process of the control system, three outputs of the output layer that they express the degree of turbidity, PH value of the water and the flocky precipitate were connected to three sensors. The flocculation agent, alkali and chlorine neurons of the input layer were connected to three dosages controller of flocculation agent, alkali and chlorine respectively. Water temperature of input layer linked the controller of temperature. If the turbidity of water, PH value or the flocky precipitate was changing, three related sensors detected the change timely and inputted them to WN1 network, which could compute the dosage of the flocculation agent, alkali, chlorine and the control value of temperature through the network compute. After this, if the quality of water also could not satisfy the anticipate effect, the WN1 network could be trained on-line using the error between the output of WN1 and the anticipant output until get the proper dosage. That is mean to control the dosage real-time through several controllers.

4. APPLICATION

Before the system work, we collected the training samples. We could obtain large numbers of relations between the dosage of the flocculation agent, alkali, chlorine, the temperature of water and the degree of turbidity, PH value of the water, the flocky precipitate by experiment, then we could get many group training samples. Every training sample was composed of 4-inputs and 3-outputs. We ascertained the number of the basic function of the wavelet neural network according to the sparseness of the sample data and ascertained the initial value of the others every parameter of network. When the structure and every parameter initial value of the wavelet neural network were ascertained, using the learning algorithm to train the network in order to obtain every parameter value. The network learning was divided into off-line learning and on-line learning two phases. In the off-line learning phase, the wavelet network was trained using a group of training sample, the parameters that trained include the network weights w_{jk}, a_j, b_j and r_j. In the on-line learning phase, consider the calculated amount and the real time property problem, only adjust the network weights w_{jk}.

In the working process, first, the controlled object identification model WN2 was trained off-line. Then we used the trained WN2 to train the feed-forward controller WN1. After WN1 was confirmed, the system could work on-line and also the WN1 could be trained on-line. In the system control process, only the feed-forward controller WN1 takes effect.

We write the corresponding control program using C++. In order to show the advantage and feasibility of genetic wavelet neural network, we adopted genetic wavelet neural network and BP network to control the flocculation process real-time. The structure of BP was the same as the genetic wavelet neural network, namely BP network also adopted the three layer structure, and had one hidden layer. The input layer of the wavelet neural network and the BP network had the same number of input neurons, that was 4, and the output layer had the same three output neurons. According to the actual sample data and the programme computing, the number of the basic function of wavelet neural network was 8, then the hidden layer of BP network had 8 neurons also. So the topology structure both of wavelet neural network and the BP network was 4-8-3. The learning parameters of network were selected as follow: the learning efficiency was 0.7, the inertia coefficient was 0.9, the system maximum error was 0.01, the maximum error of single sample was 0.001, the iteration times of network was 5000.

We could random separate the 500 samples into 5 groups, and take out 80 sample data as training samples each time, while the other 20 samples as testing sample. After the normalized process of the sample set, the genetic wavelet neural network and BP network was trained 5 times respectively using 5 groups different training sample, thenceforth we used the corresponding testing samples to test. The training error and testing error were respectively the average value of the 5 times training error and 5 times testing error. The result of train and test by genetic wavelet neural network and BP network shows as table 1.

Table 1. The training and testing results by genetic wavelet neural network and BP network

Method	Training error	Testing error	Average iteration times
Genetic wavelet neural network	0.049	0.087	250
BP network	0.083	0.175	900

From the table 1, we can see the mean squared error of training samples of genetic wavelet neural network is smaller than that of BP network. Moreover the mean squared error of testing samples of genetic wavelet neural network is also smaller than that of BP network. Furthermore the iteration times of the genetic wavelet neural network is obvious smaller than that of BP network with the same system error. It shows that the genetic wavelet neural network is all superior to the BP network at the estimation accuracy, prediction accuracy and convergence rate aspects. So the control system based on the genetic wavelet neural network has higher stability and faster real-time controlled speed.

5. CONCLUSION

Because the wavelet neural network has the advantages of both wavelet analysis and neural network, so it has the faster convergence rate and the stronger approaching ability in relation to the BP network. Meanwhile the genetic algorithm has the global search ability. Therefore the intelligent control system based on the genetic wavelet neural network in this paper can enhance control accuracy and control speed of flocculation process to a great extent. The control system is tested by experiment data, and the testing precision of the system is verified through the verification method of circular sample. The results show that the system can control the flocculation process of sewage treatment real-time and effectively and conveniently. The flocculation intelligent control technique has a very wide long term potential in the sewage treatment.

REFERENCES

1. Ren Min, Wang Wanliang, Li Tanwei, Guan Qiu, Yao Minghai. The Study on Neural Network Based Soft Measuring System for Sewage Handling. Process Automation Instrumentation, 2001, 22 (10): 8-9.
2. Wu Daoji et al., Study of Kinetics Mechanism of flocculation and controlling Index, ENVIRONMENTAL ENGINEERING, 2000, 05 pp 22-25.
3. Shen H, Iha H, Yaguchi S, Tsurugi K, The mechanism by which overexpression of Gts1p induces flocculation in a FLO8-inactive strain of the yeast Saccharomyces cerevisiae. FEMS Yeast Res, 2006 Sep; 6 (6): pp.914-923.
4. Kaseamchochoung, Chudapak et al., Influence of Chitosan Characteristics and Environmental Conditions on Flocculation of Anaerobic Sludge, Water Environment Research, Volume 78, Number 11, November 2006, pp. 2210-2216.
5. Kobayashi O, Suda H, Ohtani T, Sone H, Molecular cloning and analysis of the dominant flocculation gene FLO8 from Saccharomyces cerevisiae. Mol Gen Genet, 1996 Jul 26; 251(6): 707-715.
6. XIANG Zhuang-li. On flocculation control technology [J]. Engineering design and construction, 2004, 36 (3):40-42.
7. Quan Jiping, Huang Xiaodong, Xiao Weigui, et al., Comparison between FCD and SCD two automic feeding systems [J]. Industrial Water Treatment, 2005, 25(2): 72-74.
8. Special Issue on Fuzzy Logic and Neural Networks. IEEE Transactions on Neural Networks, 1992, 3 (5): 68-84.
9. Jiao Licheng. Neural network system theory [M]. Xian: Xi an electronic science and technology university press, 1995.
10. Pati Y C, Krishnaprasad P S. Analysis and sythesis of feedforward neural networks using discrete affine wavelet transformations. IEEE Trans Neural Networks, 1993, 4 (1):73-85.
11. Zhang Q, Benveniste A. Wavelet networks. IEEE Trans Neural Networks, 1992, 3(6): 889-898.
12. Zhao Zhenning, Xu Yongmao. Introduction to fuzzy theory and neural networks and their application [M], Beijing: Tsinghua University Press, 1996.
13. Grefenstetle J. Optimization of control parameters for genetic algorithm [J]. IEEE Trans on Syst, Man and Cybern, 1986,16(1):122-128.
14. Huang Xiuxuan, Zhu Xuefeng. Modified adaptive genetic algorithm [J]. Academic Periodical Abstracts of China, 1998,14(11): 1415-1417.
15. Li Shiyong. Fuzzy Control Neurocontrol and Intelligent Cybernetics. Harbin: Harbin Institute of Technology publishing house, 1996.
16. Xie Meiping. The non-linear time series approximation based wavelet neural network. Master's degree paper of Harbin Engineering University, 1998.