doi:10.21311/001.39.10.43

Ecological Model of Groundwater Environment Based on Hybrid Soft Computing Method

Zuihai Li *
School of Continuing Education, Yulin University, Yulin 719000, Shaanxi, China
*Corresponding author(E-mail: lizuihai@163.com)

Feng Zhang
School of Information Engineering, Yulin University, Yulin 719000, Shaanxi, China

Abstract
For mining groundwater ecosystems highly complex and non-linear, an ecological model of groundwater environment based on hybrid soft computing method (GE_HSC) was established from the point of view of algorithm optimization, and it’s includes the difference evolution algorithm and the neural network module. The experimental results show that there is no significant positive correlation between the correlation of the samples and the prediction accuracy, while the network structure and the representation of the samples have an important influence, and the water resources, environment and economy in Yulin are in uncoordinated stage at present, the economic development is faster, but the water resources are relatively poor, and economics development of the environment caused great destruction, mining area water pollution is more serious.

Key words: Hydroinformatic, Water Ecological, Hybrid Soft Computing, BP Neural Network.

1. INTRODUCTION
Groundwater resources in Yushen mining area, coal resources are relatively scarce, unreasonable exploitation caused enormous damage to groundwater resources, a serious threat to the sustainable development of the area, dynamic regulation of water resources is an effective way to alleviate the problem. The area is located in the Loess Plateau of Northern Shaanxi and Maowusu Desert border areas, lack of water resources, a fragile ecological environment, the exploitation of coal resources, large-scale destruction of the ecological environment of groundwater resources and coal resources in an integrated system, put forward the urgent demand for groundwater protection and control (Zhang and Wang, 2014). But so far, no protection and regulation of a comprehensive dynamic control model can support the groundwater resources effectively, greatly restricts the mining area groundwater resources protection, to achieve a breakthrough in the basic theory of related aspects.

Groundwater is an important part of water resources circulation system; groundwater dynamics is not only affected by the hydrogeological conditions of aquifer itself, but also affected by the external environment. In recent years, more and more scholars use soft computing methods to study the complex water environment simulation problems (Bao and Fang, 2007). Soft computing method does not require prior knowledge, or qualitative empirical knowledge as a reference, from a large number of monitoring data, access to the dynamics of the system law, so as to make up for the human understanding of the evolution of ecosystems on the deficiencies (Pinho and Ferreira, 2015). At the method level, soft computing mainly includes artificial neural network, fuzzy mathematics, genetic algorithm, support vector machine and so on. By using hybrid soft computing method, the knowledge, technology and method of different sources can be combined to adjust the structure and parameters of the system to adapt to the changing environment.

To accurately evaluate the groundwater recharge and recharge resources is the analysis of the hydrological cycle law and reasonable basis for water resources planning and sustainable groundwater exploitation plan (Li and Zhang, 2016); it has very important strategic significance for the sustainable development of the local economy and society, especially in arid and semi-arid water shortage area. Therefore, choose the high credibility of the simulation method to evaluate the groundwater recharge, to reveal the temporal and spatial variation of groundwater recharge and improve the theory and method of groundwater resources evaluation, has important theoretical and practical significance. This paper studies some soft computing methods such as artificial neural network, fuzzy mathematics, genetic algorithm and support vector machine, and then study the mixed form of them, a simulation model of groundwater environment hybrid soft computing method based on the established, then the model is applied to the ecological environment of groundwater simulation in Yushen mining area.

2. THEORY AND METHOD OF HYBRID SOFT COMPUTING
Soft-computing is proposed by Zadeh, founder of Fuzzy Set Theory in 1994, whose guiding principle is to develop tolerance techniques for uncertainty, imprecision and incomplete authenticity to obtain easy processing,
high robustness, Low cost and good integration of the actual method (Li, 2015). Soft computing is for hard computing, the traditional computing (hard computing) is the main features of strict, accurate and accurate. But hard computing is not suitable for dealing with many problems in real life, and soft computing can achieve low-cost solutions and robustness. At the method level, soft computing mainly includes fuzzy mathematics, artificial neural network, genetic algorithm, support vector machine and so on.

Soft computing is a very wide area, the model mainly includes cellular automata, individual based and box based, and these modes are mainly discrete in the system variables, time domain and spatial domain compared with the traditional ensemble model (Ian and Inyang, 2010), we take individual or spatial units as objects, to study their temporal evolution and spatial movement in order to obtain the spatial and temporal pattern of the system. The representative of soft computing includes artificial neural network, fuzzy mathematics, genetic algorithm, genetic programming, and chaos theory and law method. Compared with the usual conceptual methods, these methods take qualitative empirical knowledge as reference and obtain the dynamic laws of the system from a large number of monitoring data, thus making up for the deficiency of human understanding of the evolution mechanism of ecosystem.

Artificial neural network (ANN) is a branch of soft computing, (M-P model) proposed by Mc Culloch and Pitts in 1943 (Booth, 1986), which is a mathematical model of formal neuron structure. Neurons are the cells in the nervous system that are directly involved in information reception or generation, transmission and processing of information. The intelligence of the human brain is realized by a complex network of information flowing through neurons. The basic unit of operation of artificial neural network is a neuron, each neuron has a specific internal function operation, under normal circumstances, the neuron is a multi-input, single output nonlinear component, \( p_i \) is the input of neurons, the intensity of the information impact is measured by the joint weight \( \omega_i \), we can through the activation of the neuron function \( f \) to get the output value, the final output of the neuron is determined by comparing it with the threshold value in the output value. The relationship between neuron input and output is described by equation (1):

\[
y(t) = f \left( \sum_{i=1}^{n} \omega_i \cdot p_i(t) - \theta \right)
\]

Where \( y(t) \) the neuron is output; \( f(\cdot) \) is the neuron activation function; \( \omega_i \) is the joint weight; \( p_i \) is the input information; \( \theta \) is threshold.

Differential Evolution (DE) (Rahi, 2011) is a relatively new soft computing evolutionary algorithm; its overall structure is similar to genetic algorithm (GA). The main difference between genetic algorithms and evolutionary is operations difference. Evolutionary algorithm evolution process is as follows:

1) Constructing the difference degree vector \( D_{abcd} = (x_a - x_b) + (x_c - x_d), (x_a, x_b, x_c, x_d) \) is a randomly selected four individuals in the population, using \( D_{abcd} \) to impose noise on optimal individuals \( x_g \) or random individuals, resulting in a variant \( x_p = x_g + F \cdot D_{abcd} \).

2) Selecting the cross-factor \( R \) , a random integer \( rmb(i) \) on \([1,n]\) is generated for each \( x_i = (x_{i1}, x_{i2}, ..., x_{iu}) \) of a population, and then \( x_i \) and \( x_p \) intersect to produce a new individual \( v_i \).

\[
v_i(t+1) = \begin{cases} x_{ij}, & \text{if} \ (rand(j) < R) \\ x_{ij}, & \text{if} \ (rand(j) \geq R) \end{cases}
\]

Progeny \( v_j \) generated by \( x_i \) and \( x_p \) compete with the parent for individual \( x_i \), accepting good individuals and replacing the poorer individuals with the same population size.

3. WATER ENVIRONMENT PREDICTION AND EVALUATION

Neural network and evolutionary computation methods, have a strong adaptive, self-learning ability and search capabilities, but each has its own shortcomings, Neural network using gradient descent method, good at local optimal solution search, evolutionary algorithm using random methods, good at global optimal solution search, the combination of these two types of applications become a trend. The combination of different evolutionary algorithms and neural networks has attracted more and more attentions. The research in this field is very active and has made a lot of achievements. It has brought a prospect for the evolutionary computation and the application of neural network. In this paper, DE-BP artificial neural network is established by using DE algorithm and neural network (Quiroga and Popescu, 2013). DE algorithm searches the initial joint weights from the global point of view for the neural network, and the neural network continues to search for the optimal solution using the BP algorithm until the required solution is found. DE-BP algorithm mainly uses the global
search capability of DE algorithm (Bagirov and Barton, 2013), avoids the "premature" phenomenon in BP algorithm training process and falls into the local optimal solution, and optimizes the neural network structure to design the neural network with the optimal topology structure.

The basic BP algorithm consists of two aspects (Yuan Xu and Huifeng Xue, 2016): the forward propagation of the signal and the inverse propagation of the error, i.e., the calculation of the actual output in the direction from input to output, while the weight and threshold of the correction from the output to the input direction, as is shown in Figure 1.

![DE-BP network structure diagram](image)

**Figure 1.** DE-BP network structure diagram

BP algorithm is one of the most mature training algorithms for neural network training because of its simple, easy to do, small amount of computation and strong parallelism. However, BP has low learning efficiency, slow convergence speed and easy to fall into local minimum state. Due to the lack of historical data, the collected data samples are limited. Based on the shortcomings of BP neural network and the limited environmental sample data in Yulin area, BP neural network and DE algorithm are used to optimize the learning of BP network, which can accelerate the convergence and avoid the local minimum Has certain effect, and its change tendency forecast.

In Figure 1, the input \( net_i \) of the \( i \) node of the hidden layer is calculated as follows:

\[
net_i = \sum_{j=1}^{M} v_j x_j
\]

(3)

The output \( y_i \) of the \( i \) node of the hidden layer is calculated as follows:

\[
y_i = f(net_i) = f\left(\sum_{j=1}^{M} v_j x_j\right)
\]

(4)

The input \( net_k \) of the \( k \) node of the output layer is calculated as follows:

\[
net_k = \sum_{i=1}^{q} w_{ik} y_i = \sum_{i=1}^{q} w_{ik} f\left(\sum_{j=1}^{M} v_j x_j\right)
\]

(5)

The output \( o_k \) of the \( k \) node of the output layer is calculated as follows:

\[
o_k = f(net_k) = f\left(\sum_{i=1}^{q} w_{ik} f\left(\sum_{j=1}^{M} v_j x_j\right)\right)
\]

(6)

In this paper, a set of data \((t_0, f_0), (t_1, f_1), \ldots, (t_n, f_n)\) for a certain environmental index \( x \) is established as follows: the function approximation neural network model with single hidden layer:

\[
x(t, D) = \sum_{i=0}^{\infty} \omega_i e^{\left(\frac{t_i - b_i}{a_i} - y_i\right)}
\]

(7)
Let $D$ be a positive real variable, for $i = 0, 1, ..., n$, set $a_i = 1$, $\alpha(t) = \left(\frac{3\pi^2}{\sqrt{5}}\right)(1-t)^\frac{1}{2}$.

$x = 0.5(x_u + 1)(x_{\text{max}} - x_{\text{min}}) + x_{\text{min}}$.

4. SIMULATION AND APPLICATION FOR WATER ECOLOGICAL ENVIRONMENT

Groundwater ecosystem is a complex system with multi-factor coupling, the relationship between ecological factors is complex, showing great randomness, uncertainty and nonlinearity (Wen-Sheng and Hui-Feng, 2014). The interaction between the various factors in the system and its dynamic change process are not fully known, which restricts the development of deterministic ecological hydrodynamics. Based on the data optimization, this paper establishes a simulation model of groundwater environment based on hybrid soft computing method (GE_HSC).

The basic framework of the GE_HSC model is shown in Figure 2.

![Figure 2. GE_HSC model framework](image)

The model is a multi-input single-output model, usually the model input is a number of groundwater environmental indicators, and the model output is the water quality index data. The role of neural network and support vector machine in the model is similar, the relationship is tied, and we can choose one of them.

In this paper, a monitoring area of groundwater in Yulin city was chosen as the object of study. The data of the model establishment and verification were measured by the measured data. Yulin city has 12 watershed zoning, in order to measure the synchronous series calculation of the average total of 19.46 billion m$^3$ river runoff, and the annual runoff of 50%, 75% and 95% is 18.47, 15.04 and 11.36 billion m$^3$ respectively. Among them, the total annual runoff of 10 outflow water system areas is 18,099 billion m$^3$.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Watershed area (km$^2$)</th>
<th>Average annual runoff (Billion m$^3$)</th>
<th>Average annual runoff depth (mm)</th>
<th>Different frequency annual runoff (Billion m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
</tr>
<tr>
<td>Huang Fuchuan</td>
<td>233</td>
<td>1.032</td>
<td>48.55</td>
<td>1.098</td>
</tr>
<tr>
<td>Shimizugawa</td>
<td>421</td>
<td>0.234</td>
<td>51.23</td>
<td>0.212</td>
</tr>
<tr>
<td>Gushanchuan</td>
<td>312</td>
<td>0.215</td>
<td>78.33</td>
<td>0.421</td>
</tr>
<tr>
<td>Kuye river</td>
<td>5123</td>
<td>3.126</td>
<td>60.12</td>
<td>2.312</td>
</tr>
<tr>
<td>Wuding River</td>
<td>8986</td>
<td>5.543</td>
<td>43.30</td>
<td>4.234</td>
</tr>
<tr>
<td>Total</td>
<td>15075</td>
<td>10.15</td>
<td>281.53</td>
<td>8.182</td>
</tr>
</tbody>
</table>

Table 1. Immigration river runoff calculation

Through the analysis in Table 2, we can see that the chemical type of diving water in the area is relatively simple, and the water chemistry types of desert beach and loess hilly region are generally HCO$_3$-CaMg type, Plains in some areas, due to strong concentration, the water type is mostly ClSO$_4$-N$_6$-type water.
In order to test the effectiveness of the model, the sample data set is divided into two parts: one is the training sample set, which is used to build the model; the other part is the test sample set, which is used to verify the model. The test sample does not participate in the process of model establishment. The monitoring data of the mining area is taken as the test sample set, and the model is validated. The monitoring data of the remaining stations are training sample data.

<table>
<thead>
<tr>
<th>River</th>
<th>PH value</th>
<th>Dissolved oxygen (mg/L)</th>
<th>Oxygen consumption (mg/L)</th>
<th>Degree of mineralization (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luhe (Hengshan)</td>
<td>7.3</td>
<td>2.9</td>
<td>5.7</td>
<td>0.87</td>
</tr>
<tr>
<td>Wuding River (Xiangshui)</td>
<td>7.5</td>
<td>3.4</td>
<td>2.5</td>
<td>0.69</td>
</tr>
<tr>
<td>Wuding River (Bai Jiachuan)</td>
<td>7.9</td>
<td>5.9</td>
<td>2.1</td>
<td>0.59</td>
</tr>
<tr>
<td>Tuweihe (upstream)</td>
<td>7.8</td>
<td>5.1</td>
<td>2.3</td>
<td>0.19</td>
</tr>
<tr>
<td>Tuweihe (downstream)</td>
<td>7.7</td>
<td>3.6</td>
<td>3.2</td>
<td>0.17</td>
</tr>
<tr>
<td>Kuyehe (upstream)</td>
<td>7.3</td>
<td>3.4</td>
<td>3.5</td>
<td>0.32</td>
</tr>
<tr>
<td>Kuyehe (middle)</td>
<td>7.8</td>
<td>5.8</td>
<td>3.9</td>
<td>0.36</td>
</tr>
<tr>
<td>Kuyehe (downstream)</td>
<td>7.6</td>
<td>4.9</td>
<td>5.1</td>
<td>0.31</td>
</tr>
<tr>
<td>Jia Lu River</td>
<td>7.8</td>
<td>4.1</td>
<td>2.9</td>
<td>0.26</td>
</tr>
<tr>
<td>Eight River</td>
<td>7.2</td>
<td>5.2</td>
<td>3.5</td>
<td>0.35</td>
</tr>
</tbody>
</table>

In order to process the redundant information in the original data and extract the more useful information, the data is smoothed to eliminate the singular value. Second, in order to eliminate the unit data of each index data differences, and the magnitude of the difference, to prevent the large number of "eat" decimal, the original data normalized. The input and output variables of the sample are normalized by the formula (8) to fall into the (0, 1) interval.

\[
new\_p_i = \frac{p_i - \min(p)}{\max(p) - \min(p)}, \quad 0 < i < n
\]  

(8)

Where \( p_i \) is a column vector of length \( n \) and represents one of the input factors, it is possible to increase the number of hidden layers to complicate the network and increase the training time and the tendency of "overfitting". It will reduce the result of neural network; therefore, this work uses an input layer, a hidden layer and an output layer of BP neural network.

In order to test the performance of GE_HSC neural network, the training data of groundwater level in Yulin City is used to train the network and forecast. BP network uses an intermediate hidden layer, that is, input - hidden layer - the output structure.

(1) Training parameters: network training using trial and error method (trial and error) to determine network parameters. The BP network adopts 3-5-1 structure, namely the input layer 3 nodes, the middle hidden layer 5 nodes, the output layer 1 node; hidden layer and output layer conversion using sigmoid function; the regularization of training input sample parameters is in (0.2, 0.9), the number of neurons in the hidden layer is determined by taking a numerical range as the number of hidden layer neurons, according to the previous results of the range of \((3,11,15,21,31,33,38,44,52,59,65,70,73)\), the training error function uses the training sample mean variance (V), value is 0.0001; the experimental platform uses Matlab, BP training function trainlm; the maximum number of training epochs = 500.

Write GE_HSC algorithm code on Matlab; initial population use random function in (-5, 5) interval generation, population number 500; DE algorithm scaling factor F=0.3, crossover factor R=0.2; max evolution algebra gen=600; fitness is fit = 1/V.

Sample data a total of ten years of monitoring records, of which six years of data training network, the use of one hundred thousand data for testing. Global search through GE_HSC algorithm to achieve maximum evolution algebra, \( V=0.003136621 \), the evolution process shown in Figure 3.
As can be seen from Figure 3, the representative of the selected training samples is very important, if the test sample value exceeds the range of training samples, the error increases significantly. In the training, the sample values are larger or smaller samples remain in the prediction sample, observe the prediction results found that this large or small sample error is significantly larger. In other words, training samples can cover the range of prediction samples; the overall accuracy and individual accuracy are relatively high. When the population is 200, the hidden nodes is 20, two groups of training, individual maximum relative error reaches more than 9%; the number of population is 500, two groups in the training group, the maximum error is 3.5%, while the other group was 2.1%.

5. CONCLUSIONS

In this paper, the principles and methods of soft computing methods such as fuzzy mathematics, artificial neural network, genetic algorithm and DE algorithm are studied. Combining the difference evolution algorithm (DE) with the artificial neural network (ANN), the method of determining the weights of the artificial neural network is improved. Based on the prediction of urban groundwater level and water quality, the structure and influencing factors of GE_HSC network are analyzed, the results show that there is no significant positive correlation between the correlation of the samples and the prediction accuracy, but the network structure and the representation of the samples have important influence. The results show that the correlation between the training samples and the predicted samples has no significant positive correlation.

ACKNOWLEDGEMENTS

This work is partially supported by the Funding Project for Department of Yulin University (16GK24), Natural Science and Technology Project Plan in Yulin of China (2014cxy-11) Funding Project for Department of Education of Shaanxi Province of China (15JK1861), Natural Science Basic Research Plan in Shaanxi Province of China (2016KJXX-62). Thanks for the help.

REFERENCES


