Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.12785/amisl/020305

Multi-Level Forecasting Model of Coal Mine Water Inrush based on Self-Adaptive Evolutionary Extreme Learning Machine

Zuopeng Zhao[*] *and Mengke Hu*[*]

China University of Mining & Technology, Xuzhou, China

Received: 7 Jun. 2012, Revised: 21 Sep. 2012, Accepted: 23 Sep. 2012
Published online: 1 Sep. 2014

Abstract: This paper constructs a forecasting model for coal mining water inrush from the floor through the analysis and careful study of the mechanism of water inrush in coal mining and the Self-Adaptive Evolutionary Extreme Learning Machine (*SaE-ELM*) which obtains self-learning, generalization performance and speediness is used. In SaE-ELM, the network hidden node parameters are optimized by the self-adaptive differential evolution algorithm, whose trial vector generation strategies and their associated control parameters are self-adapted in a strategy pool by learning from their previous experiences in generating promising solutions. Large amounts of historical data of mining water inrush is collected and the main controlling factors are extracted as sample data to train and test the forecasting model by the SaE-ELM, which can forecast both the existence of a water inrush from the floor and the level of the water inrush. Experiments are given to prove that the proposed method reduces the time of model construction and computation, and improves the speed and accuracy of the forecast of coal mining water inrush.

Keywords: Coal Mining Water Inrush, Forecasting Model, Extreme Learning Machine (*ELM*), Self-Adaptive Evolutionary Extreme Learning Machine(*SaE-ELM*), Support Vector Machine (*SVM*)

1 Introduction

Coal mining water inrush is one of the five big disasters in coal mining, so a quick and accurate forecast of water inrush is a bulwark for the safety in production of the coal mining [1]. Water inrush forecast refers to hydrogeological conditions, rock mechanics, mining conditions and other factors. The complex nonlinear relationship among the above factors makes it difficult to use the traditional mathematical theory to contrast the forecasting model.

Back-propagation (*BP*) algorithm, support vector machine (*SVM*) algorithm and many other algorithms have been applied to the coal mine water inrush forecast with the development of computational intelligence.Wang [2] established the forecasting model of coal mining water inrush based on BP neural network which is trained by the genetic algorithm. This method improves the accuracy of training, however, BP neural network needs a lot of time for parameter adjustment because of the structural characteristics of BP neural network. Qin [3] used the support vector machine to establish the forecasting model to forecast and prevent water disaster in coal mine, whose radial basis function and parameter were gained by grid search and 5-folder cross validation. Although this method achieved good effect,it consumes a lot of time to adjust the parameters of SVM.

Recently, a new method called extreme learning machine (*ELM*) [4,5] was developed for SLFNs and was popular for its fast training speed by means of utilizing random hidden node parameters and calculating the output weights with least square algorithm [6,7,8]. These features enable ELM to overcome several limitations existed in gradient-descent based algorithms, such as stuck in the local minima and slow convergence performance [9]. Zhao [10] used principal component analysis (*PCA*) combined with ELM to establish forecasting model, whose speed and accuracy are enhanced, but for the hidden node parameters are randomly chosen and they remain unchanged during the training phase, the model trained by ELM is random. In self-adaptive evolutionary extreme learning machine (*SaE-ELM*), the network hidden node parameters are

[*] Corresponding author e-mail: zzpeng@cumt.edu.cn,humengkeaijia@126.com

optimized by the self-adaptive differential evolution algorithm, whose trial vector generation strategies and their associated control parameters are self-adapted in a strategy pool by learning from their previous experiences in generating promising solutions [9].

In addition, China has divided the water inrush into four types based on the maximum water inrush, but most of the existed methods only forecast whether there will be a water inrush, it is not easy to make pretreatment for different water inrush situations. In this paper, SaE-ELM algorithm is used to train and test the historical data to building model, and the data contains seven main controlling factors, such as fault throw, water pressure, the thickness of waterproof layer and so on. The proposed forecasting model can be used to forecast whether there will be a water inrush and show the type of water inrush, then give a pre-warning in appropriate level.

2 SaE-ELM

SaE-ELM is an improved algorithm which is proposed for the drawbacks of ELM. ELM [4,5,6,7,8] is a new learning algorithm for single-hidden layer feed-forward neural networks (*SLFNs*), which was proposed in 2004. SaE-ELM randomly chooses hidden node parameters and analytically determines the output weights of SLFNs. In theory, this algorithm tends to provide good generalization performance at extremely fast learning speed. However, since all the network hidden node parameters of ELM is randomly generated, the model trained by ELM will be random. In SaE-ELM algorithm, an adaptive differential evolution algorithm is used to optimize the hidden node parameters, which ensures a more compact network size than ELM and avoids the limitations of ELM. And what's more, the generalization performance will be improved.

2.1 Extreme learning machine (ELM)

For ELM is the foundation of SaE-ELM, we describe the essence of ELM in this section.

For N arbitrary distinct samples$(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{in}] \in \mathbf{R}^n$ and $\mathbf{t}_i = [t_{i1}, t_{i2}, ..., t_{im}] \in \mathbf{R}^m$, the activation function is $g(x)$ and the output of a SLFN with L hidden nodes is

$$\sum_{i=1}^{L} \beta_i g_i(\mathbf{x_j}) = \sum_{i=1}^{L} \beta_i g(\mathbf{w_i} \cdot \mathbf{x_j}, b_i) = o_i, j = 1, ..., N \quad (1)$$

where $\mathbf{w}_i \in \mathbf{R}_n$ and $b_i \in \mathbf{R}$ are the jth hidden node parameters, $\mathbf{w_i}$ is the weight vector connecting the ith hidden node and the input nodes, b_i is the threshold of the ith hidden node, and $\beta_i = [\beta_{i1}, \beta_{i1}, ..., \beta_{iL}]^T$ is the weight vector connecting the ith hidden node and the output nodes. $\mathbf{W_i} \cdot \mathbf{x_j}$ denotes the inner product of $\mathbf{w_i}$ and $\mathbf{x_j}$, $g(x)$ is activation function and Sigmoid, Sine, Hardlim and other functions are commonly used. The output nodes are chosen linear in this paper.

That standard SLFNs with L hidden nodes with activation function $g(x)$ can approximate these N samples with zero error means that

$$\sum_{j=1}^{L} \|o_j - t_j\| = 0, i.e. \quad (2)$$

there exist $\mathbf{w_i}$, β_i and b_i such that

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w_i} \cdot \mathbf{x_j} + b_i) = \mathbf{t_j}, j = 1, 2, ..., N \quad (3)$$

The above N equations can be written compactly as

$$\mathbf{H}\beta = \mathbf{T} \quad (4)$$

where

$$\begin{aligned} &\mathbf{H}[\mathbf{w_1}, ...\mathbf{w_l}, b_1, ..., b_L, \mathbf{x_1}, ...\mathbf{x_N}] \\ &= \begin{bmatrix} h\mathbf{x_1}) \\ \vdots \\ h(\mathbf{x_N}) \end{bmatrix} \\ &= \begin{bmatrix} g(\mathbf{w_1} \cdot \mathbf{x_1} + b_1) & \cdots & g(\mathbf{w_L} \cdot \mathbf{x_L} + b_L) \\ \vdots & \vdots & \vdots \\ g(\mathbf{w_1} \cdot \mathbf{x_N} + b_1) & \cdots & g(\mathbf{w_L} \cdot \mathbf{x_N} + b_L) \end{bmatrix}_{N \times L} \end{aligned} \quad (5)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad (6)$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad (7)$$

As named in Huang et al. [5,6], H is called the hidden layer output matrix of the neural network; the ith column of H is the ith hidden node output with respect to the inputs $x_1, x_2, \ldots x_N$.

Given any small positive value $\varepsilon > 0$ and activation function $g : R \to R$ which is infinitely differentiable in any interval, there exists such that for N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbf{R}^n$ and $\mathbf{t}_i \in \mathbf{R}^m$, for any w_i and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, $\|\mathbf{H}_{N \times L}\beta_{L \times m} - \mathbf{T}_{N \times m}\| < \varepsilon$.

In most cases of practical application, the number of hidden nodes L is less than N, so we can not find the parameters to make $\mathbf{H}\beta = \mathbf{T}$. In order to find specific $\mathbf{w}_i^a, b_i^a, \beta_i^a (i = 1 \ldots L)$ such that

$$\begin{aligned} &\|\mathbf{H}(w_1^a \ldots w_L^a, b_1^a \ldots b_L^a)\beta - \mathbf{T}\| \\ &= \min_{\beta} \|(w_1 \ldots w_L, b_1 \ldots b_L)\beta - \mathbf{T}\| \end{aligned} \quad (8)$$

We should find the smallest norm least-squares solution of the above linear system through formula (9)

$$\beta = \mathbf{H}^{+}T \tag{9}$$

where $\mathbf{H}^{+}$ is the Moor-Penrose generalized inverse of the hidden layer output matrix $\mathbf{H}$.

2.2 Parameter Optimization

In SaE-ELM, the network hidden node parameters are optimized by the self-adaptive differential evolution algorithm. Differential Evolution (*DE*) is arguably one of the most powerful stochastic real-parameter optimization algorithms in current use. It is a simple yet efficient evolutionary technique for many real-world optimization problems [11]. It through three operations including mutation, crossover and selection to produce the new population, and the process is repeated until the stop condition is met. Its success is highly dependent on the choice of correct trial vector generation strategies and control parameters [12]. However, the user needs to find the best values for the control parameters of DE for each problem. Finding the best values for the control parameters for each problem is a time consuming task [13]. In the adaptive differential evolution algorithm, a set of control parameters F and crossover rate CR are randomly generated for each target vector. The mutation strategy was chosen through the evaluation of fitness function.

A set of NP vectors where each one includes all the network hidden node parameters are initialized as the populations of the first generation

$$\theta_{k,G} = [\mathbf{w}^T_{1,(k,G)}, \ldots, \mathbf{w}^T_{L,(k,G)}, b^T_{1,(k,G)}, \ldots, b^T_{L,(k,G)}] \tag{10}$$

where $\mathbf{w}_j$ and $b_j\, j = (1, \ldots, L)$ are randomly generated, G represents the generation and $k = 1, 2, \ldots, NP$.

Then DE employs a self-organizing scheme that takes the difference vector of randomly chosen population vectors to perturb an existing vector to generate a new mutant vector. Here, we list four mutation strategies that we will use as follows:

$$\begin{aligned}
\mathbf{v}_{k,G} &= \theta_{r1,G} + F(\theta_{r2,G} - \theta_{r3,G}) \\
\mathbf{v}_{k,G} &= \theta_{r1,G} + F(\theta_{best,G} - \theta_{r1,G}) \\
&\quad + F(\theta_{r2,G} - \theta_{r3,G}) + F(\theta_{r4,G} - \theta_{r5,G}) \\
\mathbf{v}_{k,G} &= \theta_{r1,G} + F(\theta_{r2,G} - \theta_{r3,G}) + F(\theta_{r4,G} - \theta_{r5,G}) \\
\mathbf{v}_{k,G} &= \theta_{r1,G} + K(\theta_{best,G} - \mathbf{v}_{k,G}) + F(\theta_{r2,G} - \theta_{r3,G})
\end{aligned} \tag{11}$$

In all this equations, F is randomly generated according to the normal distributions $N(0.5, 0.3)$, the indices $r1, r2$, $r3, r4, r5$ are mutually exclusive integers randomly generated within the range $[1, 2, \ldots NP]$, which are also different from the index i. The control parameter K is randomly generated within the region $0 \le K \le 1$.

After generating all the mutant vectors, a crossover procedure is used to increase the diversities of the perturbed parameter vectors. A trial vector $u_{k,G}(j)$ is created according to the following crossover equation

$$u_{k,G}(j) = \begin{cases} v_{k,G}(j) & if(rand_j \le CR) or (j = j_{rand}) \\ \theta_{k,G}(j) & otherwise \end{cases} \tag{12}$$

where CR is the crossover rate to control the fraction of the parameter values copied from the mutant vector and is randomly generated according to the normal distributions $N(0.5, 0.1)$. $rand_j$ is the jth evaluation of a uniform random number generator with outcome in [0, 1]. j_{rand} is a randomly chosen integer from $[1, L]$ and is introduced to ensure that there exist at least one parameter in $u_{k,G}$ differing from the target vector $\theta_{k,G}$.

Calculate the network output weight matrix and root mean square error (*RMSE*) with respect to each population vector with equation (9) and the following equation, respectively.

$$\begin{aligned}
&RMSE_{k,G} \\
&= \sqrt{\frac{\sum_{i=1}^{n} \| \sum_{j=1}^{L} \beta_j g(\mathbf{w}_{j,(k,G)}, b_{j,(k,G)}, \mathbf{x}_i - \mathbf{t}_i) \|}{m \times N}}
\end{aligned} \tag{13}$$

Then use the value of RMSE to calculate the new best population vector $\theta_{k,G+1}$ with the following equation.

$$\theta_{k,G+1} = \begin{cases} u_{k,G+1} & if(RMSE_{\theta_{k,G}} - RMSE_{\theta_{k,G+1}}) \\ & > \varepsilon \cdot RMSE_{\theta_{k,G}} \\ u_{k,G+1} & if|RMSE_{\theta_{k,G}} - RMSE_{\theta_{k,G+1}}| \\ & < \varepsilon \cdot RMSE_{\theta_{k,G}} \\ & and \|\beta_{u_{k,G+1}}\| < \|\beta_{u_{k,G}}\| \\ \theta_{k,G} & otherwise \end{cases} \tag{14}$$

The three operations mutation, crossover and selection are repeated until the goal is met or the maximum iterations are reached. At last we calculate the output weights $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots \beta^T_{iL}]$ with equation (9).

3 Using the SaE-ELM to construct the coal mining water inrush forecasting model

Many methods have been used to construct forecasting model for water inrush from the floor, including SVM, BP neural network etc. But all these methods need a long time for training model, which makes it not suitable for real-time forecast. SaE-ELM can overcome the limitation of slow convergence performance which exists in gradient-descent-based algorithms, and solve the randomness consisted in ELM.

In order to construct a forecasting model for coal mining water inrush from the floor, we should analyze the mechanism of forecast for coal mining water inrush, and then choose the main factors that have influence on the coal mining water inrush. Then we need to search for plenty of historical data of main controlling factors to built sample set, and divide it into training samples and testing samples. After pre-treating the data we train forecasting model and select the appropriate activation function and the amount of hidden layer nodes in the process to ensure a fast and accurate forecasting model. A flow chart of the forecasting model of coal mine water inrush is presented in figure 1.

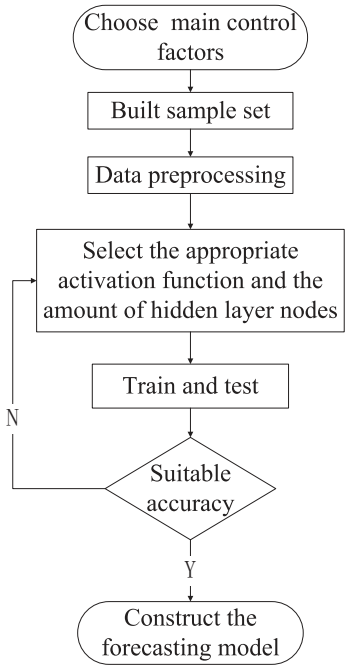


Fig. 1: Flow chart of the forecasting model of coal mine water inrush

3.1 Choose main control factors

Mining water inrush forecast refers to hydrogeological conditions, rock mechanics, mining conditions and other factors. Choosing the main factors scientifically and rationally is the precondition to construct a accurate forecasting model for water inrush from the floor. Through a great deal of research, we choose the thickness ,the water pressure and the watery of the aquifer, the thickness of waterproof layer, fault throw, mining depth and coal-mining method as the main factors.

Aquifer is the water source for water inrush, and its thickness, water pressure and the watery are the key indicators to show its influence on water inrush; Waterproof layer is the geological barrier for coal mine, and its impedance capability mainly depends on its thickness and rock properties [14]; Fault undermines the integrity of the rock bottom, and it is the main channel for the movement of groundwater and the water inrush; Mining depth is a predisposing factor of coal mining water inrush and plays a trigger role for coal mining water inrush; Coal-mining method is divided into many types, because blasting mining will increase the pressure on the floor which leads to the water inrush, we can only differentiate blasting mining and non- blasting mining [15,16].

Table 1: Water inrush types

Maximum amount of water	Water inrush type
$Q \geq 1800m^3/h$	oversize
$600m^3/h \leq Q < 1800m^3/h$	large
$60m^3/h \leq Q < 600m^3/h$	medium
$Q < 60m^3/h$	miniature

3.2 Data preprocessing

Various factors that affect water inrush should be normalized before constructing a predictive model because types of water inrush data and physical dimensions of the influencing factors are different. The principle of parameter values is that the parameter is expressed in quantitative data if the parameter can be quantitatively, otherwise expressed in binary mode. Among them, the watery of the aquifer, coal-mining method and water inrush type are expressed in binary mode. The partitions of the water inrush types are based on the maximum amount of water according to the ”mine water prevention regulations” [17], as in table 1.

The quantitative data are normalized by the following formula with outcome in [0, 1].

$$\overline{x} = (x - x_{min})/(x_{max} - x) \tag{15}$$

where x_{min}, x_{max} is the minimum and maximum of the each type date from the sample set, respectively; $x, \overline{x}$ is the values before and after normalization, respectively.

3.3 Training coal mining water inrush forecasting model by SaE-ELM

We divide the sample set which has been processed into training samples and testing samples to train coal mining water inrush forecasting model by using SaE-ELM. The model is trained in a large program which can test immediately after the training completed. According to SaE-ELM theory that has been introduced above, we can summarize the following steps.

For N arbitrary distinct samples $(x_i,t_i), i = 1,\ldots N$,and L hidden nodes and activation function $g(x)$:

Step 1: A set of NP individual parameter vectors $\theta_{k,G}(k = 1,2\ldots NP)$, where each one includes all the network hidden node parameters are initialized as the populations of the first generation;

Step 2: In the case of $g(x)$ and L are invariable run the three operations including mutation, crossover and selection to produce the new population, and the process is repeated until the stop condition is met.

Step 3: Changing the type of $g(x)$ and increase the number of hidden nodes L gradually from one to find the most suitable $g(x)$ and L to construct an optimal forecasting model with the best testing accuracy;

Step 4: Calculating the output matrix according to Eq.(4);

Step 5:Calculating the output weights $\beta : \beta = \mathbf{H}^{+}\mathbf{T}$, where $\mathbf{T} = [t_1,\ldots t_N]$ and $\mathbf{H}^{+} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$.

When the best testing accuracy has been gotten we made parameters $\mathbf{w}_i, b_i$ and β_i output, it will get the optimal coal mine water inrush forecasting model.

4 Experimental evaluation

4.1 Sample data selection

We investigate some the mining water inrush situation in China and collect the actual data of water inrush, part of the primordial sample data are shown in table (2). We just use 'Y' and 'N' to show whether the coal-mining method is blasting mining or not. Each sample contains seven inputs and one output.

The quantitative data are normalized by formula (15) with outcome in [0, 1]; The watery of the aquifer, weak, medium, strong and very strong are represented by the binary numbers 00, 01, 10, 11 respectively; The coal-mining method blasting mining or not are represented by the binary numbers 1, 0 respectively; The type of water inrush oversize, large, medium, miniature are represented by the binary numbers 110, 100, 011, 001 respectively, and the binary numbers 000 shows that there is no water inrush. The normalized sample data are shown in table 3.

4.2 Parameter Selection

All experiments on SaE-ELM, ELM and SVM are carried out in the MATLAB 7.6 environment. It has a very efficient implementation of SVM provided by Libsvm package which has been used in our simulations for SVM, the kernel function used in SVM is radial basis function, it is easy to achieve the classification when SaE-ELM is used to construct forecasting model, a set of

Table 4: Part of the primordial sample data

	Training time(s)	Accuracy(%)		Nodes/SVs
		Training	Testing	
ELM	0.0118	85.4	87.7	25
SaE-ELM	0.0254	90.2	92.6	25
SVM	0.7232	89.5	90.1	47

NP individual parameter vectors are initialized to cover the parameter space, where each vector includes all the network hidden node parameters are initialized as the populations of the first generation. Based on experience in [9], NP is set to be 20. But the forecasting performance of the model will be significantly different if the number of hidden nodes or the type of activation function is different, we need to select the number of hidden nodes under the condition of fixed activation function. We chose different activation functions (Sine function, Sigmiodal functions, Hardlim function, Triangular basis functions and Radial basis function), and for the five activation functions, the number of hidden nodes was gradually increased from 5 to 125 with the interval. Then the optimal number of nodes and the the type of activation function can be selected. The specific analysis of the results is showed in fig.2.

4.3 Choose main control factors

As observed from fig.2, the testing accuracy is always little when the activation function is hardlim. With the increase of hidden nodes, the testing accuracy of the other four functions increases at first and then decreases, so each one of these four activation functions has a maximum testing accuracy, and we can get them when the number of hidden nodes is about 25. When the number of hidden nodes is about 80, the testing accuracy reaches minimum, and then tends to be a stable surface. At last, we choose sine function as the activation function and use 25 hidden nodes to train the forecasting model.

4.4 Comparison experiments of training model by different algorithms

We use SaE-ELM, ELM and SVM trained the samples based on the optimal parameters selected in section 4.3. After 50 experiments based on the above three algorithms, we can calculate the average training time, average training accuracy and average testing accuracy to evaluate the water inrush model. We give the comparison results of the performance of the three above algorithms in Table 4.

As we can see from table 4, compared with ELM the training speed of SaE-ELM model is a little slower, but the training accuracy and testing accuracy of ELM is

Table 2: Part of the primordial sample data

coal mine	fault throw (m)	mining depth (m)	thickness of waterproof layer(m)	thickness of aquifer (m)	watery	water pressure (kpa)	coal-mining method	maximum amount of water(m^3/s)
1	8	185	55	155	strong	3.95	N	1400
2	2.5	403	53	158	very strong	3.80	N	600
3	20	60	15	60	strong	31.3	N	3153
4	20	85	35.7	55.64	strong	6.10	Y	780
5	21.2	94	59.4	69.53	weak	41.2	Y	100
6	13	40	55	130	strong	4.5	Y	220
7	0	205.2	36.94	584	strong	14.53	Y	0
8	0	105	53.62	18.24	strong	4.00	Y	200
9	1.4	90	50.00	18.42	medium	3.50	Y	40
10	0	70	15	51.13	strong	5.00	N	175

Table 3: Part of the normalized sample data

coal mine	fault throw	mining depth	thickness of waterproof layer	thickness of aquifer	watery	water pressure	coal-mining method	Water inrush type
1	0.3774	0.4453	0.7273	0.2424	10	0.0485	0	011
2	0.1179	1.0000	0.6909	0.2477	11	0.0447	0	011
3	0.9434	0.1272	0.0000	0.0746	10	0.7471	0	100
4	0.9434	0.1908	0.3764	0.0669	10	0.1034	1	011
5	1.0000	0.2137	0.8073	0.0914	00	1.0000	1	010
6	0.6132	0.0763	0.7273	0.1982	10	0.0626	1	010
7	0.0000	0.4967	0.3989	1.0000	10	0.3188	1	000
8	0.0000	0.2417	0.7022	0.0008	10	0.0498	1	010
9	0.0660	0.2036	0.6364	0.0011	01	0.0370	1	001
10	0.0000	0.1527	0.6569	0.0000	10	0.0754	0	010

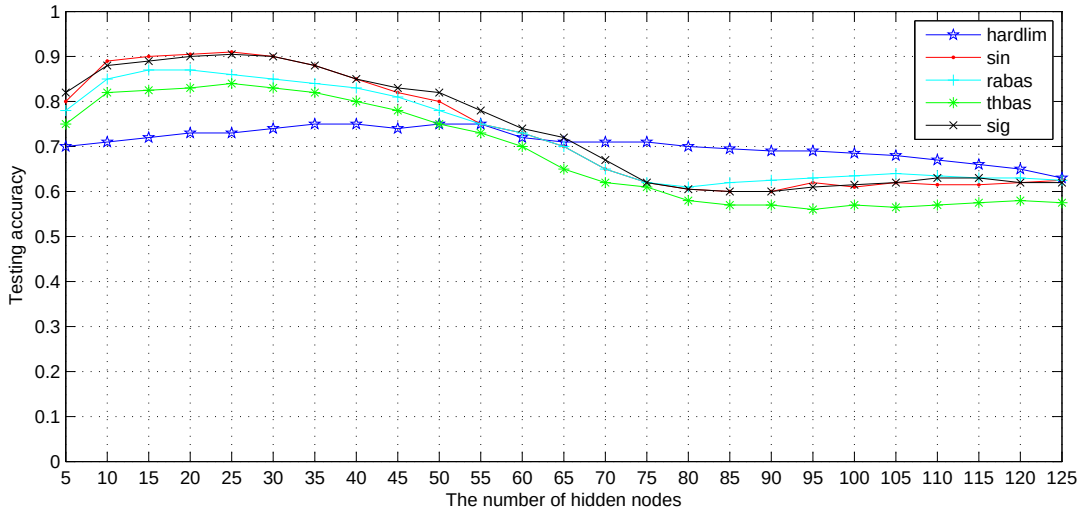


Fig. 2: The comparison chart for testing accuracy

much lower than SaE-ELM. Because that the hidden node parameters in ELM are randomly assigned and remain unchanged during the training phase, but the hidden node parameters are optimized by the self-adaptive differential evolution algorithm when SaE-ELM is used. For the optimization helps save the time spent on hidden node parameters, the accuracy is greatly improved. Thus the forecasting performance of SaE-ELM model is better than ELM.

Comparing SaE-ELM with SVM algorithm, the training time of SaE-ELM is shorter than SVM, both of the training accuracy and testing accuracy are relatively high, especially the testing accuracy of them is reached up to more than 90%, but the testing accuracy of SaE-ELM is higher than SVM even less nodes of hidden layers are used. All of above results show that the performance of ELM is better than the SVM.

5 Conclusions

It is an approximation problem of complex nonlinear function between the predicted value of mining water inrush and the factors. We get a forecasting model through training a large number of historical data by SaE-ELM. The approach proposed in this paper not only learns faster but also has a higher generalization performance and forecast accuracy than traditional algorithms. Randomness which is generated from the random hidden layer parameters in ELM is avoided. For the reason that the forecasting result of the model is the level of water inrush, we can set different corresponding warning levels to forecast the water inrush. Therefore, the method proposed in this paper can satisfy the real-time requirement of the coal mining water inrush and is of value to promote.

Acknowledgement

This research was supported by Natural Science Foundation of Jiangsu Provincial (BK2012129).Authors would also like to express their gratitude to anonymous reviewers and editor for valuable suggestions to improve this paper.

References

[1] Cui Zhu, Jiazhong Qian, Xiaoping Zhou. Water-inrush source discrimination with BP neural network in Pansan Mine. Journal of Anhui Institute of Architecture. **18**, 35-38 (2010).

[2] Lianguo Song,Yang Song. Combined ANN forecast of water-inrush from coal floor. Chinese Journal of Geotechnical Engineering. **23**, 502-505 (2001).

[3] Jiexuan Qin, Cuiping Li, Zhongxue Li. Prediction of Mine Water Inrush Quantity Based on Support Vector Regression.China Safety Science Journal.**5**, 114-119 (2013).

[4] Guang-Bin Huang, Qin-Yu Zhu, Chee-Kheong Siew. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. IEEE International Joint Conference. **2**, 985-990 (2004).

[5] Liang N-Y, Huang G-B, Saratchandran P, Sundararajan N . A fast and accurate on-line sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw .**17**,1411C1423 (2006).

[6] Huang G B, Zhu Q Y and Siew C K. Extreme Learning Machine: Theory and Applications Neurocomputing. **70**, 489-501 (2006).

[7] Huang G B, Chen L. Convex Incremental Extreme Learning Machine,Neurocomputing. **70**, 3056-3062 (2007).

[8] Huang G B, Chen L. Enhanced Random Search Based Incremental Extreme Learning Machine .Neurocomputing. **71**, 3460-3468 (2008).

[9] Jiuwen Cao, Zhiping Lin, Guang-Bin Huang. Self-Adaptive Evolutionary Extreme Learning Machine. Neural Process Lett. **36**, 285-305 (2012).

[10] Zuopeng Zhao, Pei Li, Xinzheng Xu. Forecasting Model of Coal Mine Water Inrush Based on Extreme Learning Machine. Applied Mathematics & Information Sciences. **7**,1243-1250(2013).

[11] Pravakar Roy, Md. Jahidul Islam, and Md. Monirul Islam. Self-adaptive Genetically Programmed Differential. International Conference on Electrical and Computer Engineering. **7**, 20-22 (2012).

[12] Mahamed G.H. Omran, Ayed Salman, Andries P. Engelbrecht. Self-adaptive Differential Evolution. Springer. **1**, 192-199(2005).

[13] Sudipta Das, Gopi Ram, Durbadal MandaI, et al.Self Adaptive Differential Evolution Based Thinning of Concentric Circular Antenna Arrays. International Conference on Digital Object Identifier.1-5(2013)

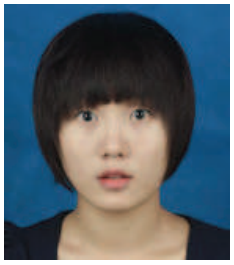
[14] Qiang Wu, Zhilong Zhang, Jifu Ma. A new practical methodology of the coal floor water bursting evaluating The master controlling index system construction.Journal China Coal Society. **32**, 42-47 (2007).

[15] Yufeng Qiao. Application Research of Genetic Algorithm and Artificial Neural Networks in the Prediction of Mine Water Gushing-out. Xi'an: Xi'an University of Architecture and Technology. 8-11 (2011).

[16] Ying LI. Prediction Method Study on Water Inrush through Coal Floor Based on Support Vetor Machines. Xi'an: China Coal Research Institute. 17-20 (2007).

[17] State Administration of Work Safety. Mine water control requirements[M]. China Coal Industry Publishing House, 7-7 (2009).

Zuopeng Zhao received the MS degree in Communication and Information System from China University of Mining & Technology in 2005, and the PhD degree in Cartography and Geographic Information Systems from the Beijing University in 2001. He is currently a vice professor in the China University of Mining Technology.His research interests are in the areas of information technology of coal mine safety, wireless sensor network (WSN),and geographic information system(GIS).



Mengke Hu received her BS degree in electric engineering and automation from China University of Mining & Technology in 2012. She is currently a master at school of Control Science and Engineering, China University of Mining Technology. Her research interests are in the areas of machine learning,intelligent information processing,and coal mine safety information technology.