Energy 122 (2017) 350–362

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Energy saving and prediction modeling of petrochemical industries: A novel ELM based on FAHP

ZhiQiang Geng [a, b], Lin Qin [a, b], YongMing Han [a, b, *], QunXiong Zhu [a, b, **]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education in China, Beijing, 100029, China

ARTICLE INFO

Article history:
Received 22 July 2016
Received in revised form
16 January 2017
Accepted 17 January 2017
Available online 23 January 2017

Keywords:
Extreme learning machine
Fuzzy C-Means algorithm
Analytic hierarchy process
Energy conservation and emissions reduction
Petrochemical industries

ABSTRACT

Extreme learning machine (ELM), which is a simple single-hidden-layer feed-forward neural network with fast implementation, has been widely applied in many engineering fields. However, it is difficult to enhance the modeling ability of extreme learning in disposing the high-dimensional noisy data. And the predictive modeling method based on the ELM integrated fuzzy C-Means integrating analytic hierarchy process (FAHP) (FAHP-ELM) is proposed. The fuzzy C-Means algorithm is used to cluster the input attributes of the high-dimensional data. The Analytic Hierarchy Process (AHP) based on the entropy weights is proposed to filter the redundant information and extracts characteristic components. Then, the fusion data is used as the input of the ELM. Compared with the back-propagation (BP) neural network and the ELM, the proposed model has better performance in terms of the speed of convergence, generalization and modeling accuracy based on University of California Irvine (UCI) benchmark datasets. Finally, the proposed method was applied to build the energy saving and predictive model of the purified terephthalic acid (PTA) solvent system and the ethylene production system. The experimental results demonstrated the validity of the proposed method. Meanwhile, it could enhance the efficiency of energy utilization and achieve energy conservation and emission reduction.

1. Introduction

Energy is an essential pillar of modern economy and the cornerstone of modern civilization. With the rapid development of economy, energy issues are increasingly prominent. From a development point of view, the energy conservation and emissions reduction now are deemed as a worldwide common themes, especially in the petrochemical industries. Therefore, the improvement of energy efficiency is an important means to achieve economic goals and solve environmental problems. The amount of the industrial ethylene production, which is the pillar of the petrochemical industry, is commonly used as one of the major indicators of a country's industry level. And its energy consumption also accounts the important all-around index for measuring the technical performance of plants. According to the report in 2014 when China National Petroleum Corporation's ethylene production amount was 10420kt/a [1], the average fuel plus power consumption (standard oil) was 571.39 kg/t for producing a ton of ethylene, which costed far more than that of world's advanced level. A huge development in energy conservation and saving is potential in ethylene fabricating industry. Meanwhile, the purified terephthalic acid (PTA), as an important chemical raw materials, also has a pivotal role in the chemical industry and closely relates to our daily life [2,3]. In China, 70% of the PTA is used to produce polyester products while 20% of them being used to produce Polyethylene terephthalate (PET), and the rest is utilized to produce the packing material. Therefore, the optimization of the PTA production process has become a hot research topic in the petrochemical industry. The relevant organizations of countries have been committed to improve the production equipment increasing energy efficiency continuously. If the production process of petrochemical industries can be accurately analyzed and predicted, the energy conservation and emission reduction of the petrochemical industry can be achieved. In order to better analyze and predict the energy efficiency of the petrochemical industry production, we propose an improved

* Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
** Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
E-mail addresses: hanym@mail.buct.edu.cn (Y. Han), zhuqx@mail.buct.edu.cn (Q. Zhu).

http://dx.doi.org/10.1016/j.energy.2017.01.091

extreme learning machine (ELM) based on fuzzy C-Means algorithm (FCM) integrating analytic hierarchy process (AHP) (FAHP-ELM).

The organization of the remainder is as follows: Section 2 presents the research status of the energy saving and prediction with the ELM and the AHP. The details of the FAHP-ELM are introduced in Section 3. Section 4 presents the comparisons with the other feed-forward neural network by University of California Irvine (UCI) benchmark datasets. Section 5 shows two practical cases study about the energy efficiency analysis and prediction of petrochemical industry based on the FAHP-ELM. Finally, the concluding remarks are given in Section 6.

2. Related work

In recent decades, as the result of the development of the computer technology, a variety of artificial intelligence methods (expert system, genetic algorithm, neural network etc.) are widely applied for the simulation of production process as well as prediction models, and have achieved good results.

The artificial neural network (ANN), adjusts the connected relationship between many internal nodes to achieve the goal of information processing without regard to the internal mechanism. Therefore, it has the characteristics of fast response and functions as complex nonlinear approximation, etc. and has been widely applied to build the prediction model of the process industries, such as wind farms system optimizing [4], global radiation forecasting [5], power industry enhancing [6], diesel engine estimating [7] and power density predicting [8], particularly in the petrochemical industry. Monedero et al. developed a decision system based on ANN to optimize the energy efficiency of a petrochemical plant [9]. Avramović et al. used the Response surface methodology (RSM) and ANN approaches for modeling the content of fatty acid ethyl esters (FAEE) and optimizing the process variables [10]. He et al. developed a novel double parallel ELM with Pearson correlation coefficient based independent subnets (PCCIS-DPELM) for accurately modeling complex chemical processes [11]. Haghbakhsh et al. presented a new approach based on ANN for the prediction of density of pure hydrocarbons [12]. He et al. proposed the Auto-associative Hierarchical Neural Network (AHNN) to explore the problems of monitoring chemical processes with large numbers of input parameters [13]. The back-propagation (BP) neural network is a kind of error back-propagation algorithms based on the gradient descent [14], which is the most common modeling algorithm of neural networks. However, the rate of convergence is slow and easily trapped in local optimum, which does not apply to complex industrial modeling. Compared to the BP neural network, the training speed of the radical basis function (RBF) neural network is much faster and has the best overall approximation performance without the local optimization problem of learning [15,16]. However, it has high requirements to typical training sample extraction, and the difficulty of its training is high [17]. In 2004, the ELM was first proposed by Huang et al. [18]. Different from the other feed-forward neural networks, the ELM does not need to constantly adjust the weight according to the error function. The learning parameters of neuron in the hidden layer are randomly generated, while the output weights are determined by calculating the Moore-Penrose generalized inverse matrix. Thus the training speed and the generalization accuracy are high, having strong robustness and not being prone to local optima [19]. Due to these advantages, the ELM has been used in self-organized clustering [20], regression and multiclass classification [21], traffic sign recognition [22], image recognition [23], computer vision processing [24] and feature selection [25]. Cao et al. used the self-adaptive differential evolution algorithm to optimizing the learning parameters of the hidden neuron and obtained an improved self-adaptive evolutionary ELM learning algorithm [26]. Lan et al. proposed a constructive enhancement for online sequential ELM (CEOS-ELM) [27], which can add random hidden nodes one-by-one or group-by-group with fixed or varying group size. Tang et al. proposed a new ELM-based hierarchical learning framework for multilayer perceptron to solve the issue of natural signals. The proposed architecture is divided into two main components: (1) self-taught feature extraction followed by supervised feature classification and (2) they are bridged by random initialized hidden weights [28]. Gastaldo et al. presented a new variant of the basic ELM by connecting the models to improve over the conventional ELM in terms of the trade-off between classification accuracy and predictor complexity [29]. Although the learning performance of the ELM has been greatly improved, there are still many shortcomings in the production prediction of petrochemical industries. Because the complexity of petrochemical industrial data has characteristics of high dimensions, strong coupling and redundant information, we cannot accurately predict their production status based on the ELM. Therefore, it is necessary to extract features of the samples for high-dimensional data to filter noise and remove redundant information.

The AHP is a kind of integrated methods of qualitative and quantitative analysis, which is systematic and hierarchical. Due to the practicality and effectiveness in dealing with multiple-attribute decision-making problems, the AHP has been widely applied to energy efficiency analysis of petrochemical industries. Han et al. proposed an effective approach using Interpretative structural model (ISM) to hierarchical analyze the main factors and basic mechanisms affecting the energy consumption [30]. Geng et al. used a new method based on kernel function and AHP to estimate the process capability of Tennessee Eastman industrial production [31]. Han et al. analyzed and evaluated the energy efficiency of petrochemical processes using the data envelopment analysis (DEA) integrating the AHP [32]. Yagmur studied the multi-criteria evaluation and priority analysis for localization equipment in a thermal power plant using the AHP [33]. However, in order to ensure the accuracy of prediction, the input of the AHP should be a dataset with the high similarity, so the data should be clustered at first. As a clustering algorithm with excellent performance, the FCM is evolved from the K-Means clustering algorithm by utilizing the fuzzy theory to find natural vague boundaries in data which has various types of inherent degradations [34,35]. Pimentel et al. proposed a multivariate version of the FCM algorithm with weighting [36]. Zainuddin et al. presented an effective FCM algorithm that adopts the symmetry similarity measure to search for the appropriate clusters, regardless of the geometric structures and overlapping characteristic [37]. Zhang et al. introduced some prior distribution information of the missing values into the algorithm of fuzzy clustering [38].

From what has been discussed above, we propose an energy saving and prediction model based on the FAHP-ELM. The FCM algorithm is used to cluster the input attributes of the high dimensional data, and then the redundant information and extracted characteristic component are filtered using the AHP based on the entropy weight. The treated data is set as the input of the ELM, to obtain the excellent learning performance of the FAHP-ELM through training. Demonstrating the effectiveness of the proposed prediction model by UCI standard data sets and the PTA solvent system and the ethylene system in petrochemical industries, the experiment shows that, compared with the BP neural network and the ELM, the FAHP-ELM has good generalization accuracy and learning efficiency to handle the complex, high-dimensional and strong coupling data. Meanwhile, the proposed method is able to detect the production level of the petrochemical industry and enhance the energy efficiency.

3. FAHP- ELM

This FAHP-ELM aims to achieve data compression by the weighted fusion of high-dimensional data samples, the extraction of the data characteristics, the removal of redundant information and the filtration of noise. With the FAHP fusion data as the input of the ELM, the established FAHP-ELM is suitable for the complex data modeling.

3.1. Extreme learning machine

The ELM mainly is composed of an input layer, a hidden layer and an output layer. The structure of the ELM is shown in Fig. 1. Different from the other feed-forward neural network, the ELM does not require to constantly adjust the weights according to the error function, its learning parameters between the input layer and the hidden layer are randomly generated, while the output weights are analytically calculated out by using the least-square method [18].

Given a training set $= \{(\boldsymbol{I_n}, \boldsymbol{T_n})|\mathrm{n} = 1, 2, \ldots N; \boldsymbol{I_n} \in \boldsymbol{R}^W; \boldsymbol{T_n} \in \boldsymbol{R}^V\}$ where $\boldsymbol{I_n}$ is the input data vector, and $\boldsymbol{T_n}$ is the target value vector of each sample. Considering there are K hidden nodes, the output of the ELM is as follows:

$$\boldsymbol{T_n} = \sum_{k=1}^{K} \beta_{\boldsymbol{k}} G(\omega_{\boldsymbol{k}} \cdot \boldsymbol{I_n} + \boldsymbol{b_k}) \tag{1}$$

where $\omega_{\mathbf{k}} = [\omega_{\mathbf{k1}}, \omega_{\mathbf{k2}}, \ldots, \omega_{\mathbf{kW}}]^{\mathrm{T}}$ and $\beta_{\boldsymbol{k}} = [\beta_{k1}, \beta_{k2}, \ldots, \beta_{kV}]^{\mathrm{T}}$ $k = 1, 2, \ldots K$ show the input weights and the output weights of the k-th hidden layer neuron, respectively. $\boldsymbol{b_k}$ and $G(\omega_{\boldsymbol{k}} \cdot \boldsymbol{I_n} + \boldsymbol{b_k})$ represent the biases and the output of the k-th hidden layer neuron.$G(\cdot)$ is the activation of the hidden layer, and it has been defined as follows:

$$G(\boldsymbol{x}) = \frac{1}{1 + e^{-(x)}} \tag{2}$$

Eq. (1) can be transformed into a compact form, as:

$$\boldsymbol{H}\beta = \mathbf{T} \tag{3}$$

where $\boldsymbol{H}$ represents the output matrix of hidden layer (randomized matrix):

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{G}(\omega_1 \cdot \boldsymbol{I}_1 + \boldsymbol{b}_1) & \cdots & \boldsymbol{G}(\omega_{\boldsymbol{K}} \cdot \boldsymbol{I}_1 + \boldsymbol{b_K}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{G}(\omega_1 \cdot \boldsymbol{I_N} + \boldsymbol{b}_1) & \cdots & \boldsymbol{G}(\omega_{\boldsymbol{K}} \cdot \boldsymbol{I_N} + \boldsymbol{b_K}) \end{bmatrix}_{N \times K}$$

$$\beta = \left[\beta_1^{\boldsymbol{T}}, \beta_2^{\boldsymbol{T}}, \ldots, \beta_{\boldsymbol{K}}^{\boldsymbol{T}}\right]^{\mathrm{T}}_{K \times V}$$

$$\boldsymbol{T} = \left[\boldsymbol{Y}_1^{\boldsymbol{T}}, \boldsymbol{Y}_2^{\boldsymbol{T}}, \ldots, \boldsymbol{Y}_{\boldsymbol{K}}^{\boldsymbol{T}}\right]^{\mathrm{T}}_{N \times V}$$

Calculate the least-square solution of Eq. (3), and then obtain the weights between the hidden layer and the output layer.

$$\widehat{\beta} = \mathbf{H}^{+}\boldsymbol{T} \tag{4}$$

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

Accordingly, given a training set$\boldsymbol{N}$, a activation of the hidden layer $G(\cdot)$ and a number of hidden layer neurons K, the ELM training algorithm can be summarized as follows:

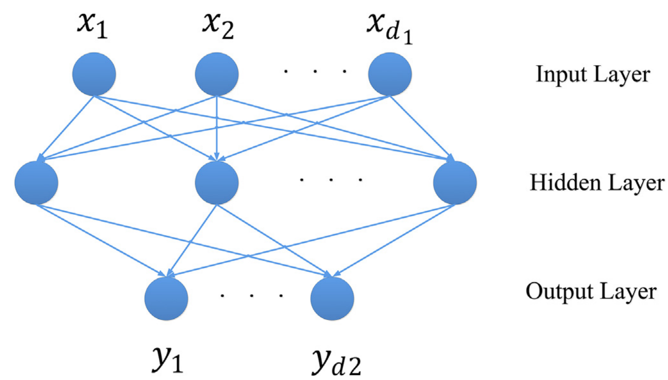

Fig. 1. The structure of the ELM.

(1) Randomly assign the input weights ω_k and the biases of the hidden layer neurons $\boldsymbol{b}_k$;
(2) Calculate the hidden layer output matrix$\mathbf{H}$;
(3) Calculate the output weights β by using the least-square solution.

3.2. FAHP

3.2.1. FCM algorithm

The FCM algorithm utilizes the fuzzy theory to find natural vague boundaries in data and uses the membership degree to determine which cluster the data belong to. Given the clustering number c and the fuzzy index p, where c is far smaller than the number of clustering samples and p is always assigned between 1.5 and 2.5 [39]. Given M samples with$\{\boldsymbol{S_m}|m = 1, 2, \ldots M; \boldsymbol{S_m} \in \boldsymbol{R}^D\}$, D is the dimension of the vector$\boldsymbol{S_m} = \{s_{m1}, s_{m2}, \ldots\ldots, s_{mD}\}$, where s_{mD} is an attribute of the sample$\boldsymbol{S_m}$. The FCM partitions D attributes into h c-fuzzy clusters $C = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots\ldots, \boldsymbol{C_h}\} \in R^M$ $\mathbf{C}_h = \{c_{1h}, c_{2h}, \ldots, c_{Mh}\}^T$ by minimizing the distance-based objective function $\mathscr{H}(U; C)$:

$$\mathscr{H}(U; C) = \sum_{i=1}^{h} \mathscr{H}_i = \left\{ \sum_{i=1}^{h} \sum_{j=1}^{D} (u_{ij})^p (Eu_{ij})^2 | \sum_{i=1}^{h} u_{ij} = 1, u_{ij} \in [0, 1], h \in [2, D] \right\} \tag{5}$$

where u_{ij} denotes the probability of the j-th attributes belonging to the i-th clustering center, U is the membership degree matrix

$$\mathbf{U} = \begin{bmatrix} u_{11} & \cdots & u_{1D} \\ \vdots & \ddots & \vdots \\ u_{h1} & \cdots & u_{hD} \end{bmatrix}_{h*D} \tag{6}$$

Then we get Eq. (7) and Eq. (8).

$$C_i = \frac{\sum_{j=1}^{D} (u_{ij})^p S_j}{\sum_{j=1}^{D} (u_{ij})^p} \tag{7}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{h} \left(\frac{Eu_{ij}}{Eu_{kj}}\right)^{2/(p-1)}} \tag{8}$$

Then, we will adopt the AHP algorithm based on the entropy weights to fusion the each cluster attributes which has been clustered by the FCM.

3.2.2. The AHP algorithm based on the entropy weights

Definition 1: Given the correlation functions of j parameters of data is$k_{ij}(x)$: (i represent the *i-th* sample)

$$k_{ij}(x)=\begin{cases}0 & x\notin[x_j(1),x_j(4)]\\ \dfrac{x_{ij}-x_j(1)}{x_j(2)-x_j(1)} & x\in[x_j(1),x_j(2)]\\ 1 & x\in[x_j(2),x_j(3)]\\ \dfrac{x_j(4)-x_{ij}}{x_j(4)-x_j(3)} & x\in[x_j(3),x_j(4)]\end{cases} \tag{9}$$

$i=1,2,\ldots,n;j=1,2,\ldots,m$

The correlation function is called the standard correlation functions, where,$x_j(1)$, $x_j(2)$, $x_j(3)$, $x_j(4)$ are nodes of $k_{ij}(x)$.

If the second node of the standard correlation function overlap in the third node,

$$k_{ij}(x)=\begin{cases}0 & x\notin[x_j(1),x_j(4)]\\ \dfrac{x_{ij}-x_j(1)}{x_j(2)-x_j(1)} & x\in[x_j(1),x_j(2)]\\ \dfrac{x_j(4)-x_{ij}}{x_j(4)-x_j(2)} & x\in[x_j(2),x_j(4)]\end{cases} \tag{10}$$

$i=1,2,\ldots,n;j=1,2,\ldots,m$

The correlation function is called the lateral correlation functions [32,40].

Assume that $\mathbf{X}=[\mathbf{X}(1),\mathbf{X}(2),\ldots,\mathbf{X}(n)]^{\mathrm{T}}$ is the data which has been clustered, and the lateral correlation function is adopted here, and $x_j(2)(j=1,2,\ldots,m)$ represent the average value. We can obtain the information matrix as follows:

$$\mathbf{K}_{n*m}=\begin{bmatrix}k_{11} & k_{12} & \cdots & k_{1m}\\ k_{21} & k_{22} & \cdots & k_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ k_{n1} & k_{n2} & \cdots & k_{nm}\end{bmatrix} \tag{11}$$

The center normalization: $k'_{ij}=(k_{ij}-\overline{k_j})/S_j \quad (i=1,2,\ldots,n;j=1,2,\ldots,m)$.

Where, $\overline{k_j}=\frac{1}{n}\sum_{i=1}^{n}k_{ij}(j=1,2,\ldots,m)$, then move the negative number to zero (and use positive zero, zero plus a positive decimal ε)$r_{ij}=k'_{ij}-t_j+\varepsilon i=1,2,\ldots,n;j=1,2,\ldots,m,$ where $t_j=\min(k'_{ij})<0(j=1,2,\ldots,m)$

Obtain the positive matrix$\mathbf{R}^{j}_{n\times m}$:

$$\mathbf{R}^{j}_{n*m}=\begin{bmatrix}r_{11} & r_{12} & \cdots & r_{1m}\\ r_{21} & r_{22} & \cdots & r_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ r_{n1} & r_{n2} & \cdots & r_{nm}\end{bmatrix} \tag{12}$$

Obtain the n-dimensional matrix by$\mathbf{R}^{j}_{n\times m}$:

$$\mathbf{COR}=\mathbf{RR}^{T}=\begin{bmatrix}o_{11} & o_{12} & \cdots & o_{1m}\\ o_{21} & o_{22} & \cdots & o_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ o_{n1} & o_{n2} & \cdots & o_{nm}\end{bmatrix} \tag{13}$$

For the n-dimensional symmetric matrix **COR**, calculated the entropy of each index: $e_i=-\frac{1}{\ln n}\sum_{j=1}^{n}(r_{ij}\ln r_{ij})(i=1,2,\cdots,n)$, where the entropy is a measure of the difference of index data in the symmetric matrix*COR*. If $r_{i1}=r_{i2}=\ldots=r_{ij}=\frac{1}{n}(j=1,2,\cdots,n)$, then e_i get the max 1. Obviously, the entropy of index data is smaller, the difference is greater, the important degree of index is greater in the comprehensive evaluation; and the entropy of index data is greater, the difference is smaller, the important degree of index is smaller in the comprehensive evaluation.

The important degree of each index is its weight:

$$w_i=\frac{1-e_i}{\sum_{i=1}^{n}(1-e_i)} \tag{14}$$

Using **W** to integrate programme, get the X_{ref}, it is the fusion data.

$$\mathbf{X}^{T}\mathbf{W}=\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1m}\\ x_{21} & x_{22} & \cdots & x_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1} & x_{n2} & \cdots & x_{nm}\end{bmatrix}^{T}\begin{bmatrix}w_1\\ w_2\\ \cdots\\ w_n\end{bmatrix} \tag{15}$$

3.3. The framework of the FAHP-ELM

Given a training sample set$\{(\boldsymbol{X_m},\boldsymbol{Y_m})|\mathrm{m}=1,2,\ldots\mathrm{M};X_m\in R^{d_1};Y_m\in R^{d_2}\}$, where $\boldsymbol{X_m}$ is the input data vector, $\boldsymbol{Y_m}$ is the target value of each sample. The details of the FAHP-ELM are provided in the following:

Step1: Cluster the high dimension input attributes of samples by using the FCM algorithm and then the each cluster has been defined as input data of the AHP algorithm;

Step2: Establish the model of the AHP by using the entropy weights. The high dimensional input vector of the training sample$\boldsymbol{X_m}=[x_{m1},x_{m2},\ldots,x_{md_1}]$, $m=1,2,\ldots M$; has been used as the input vector of the analytic hierarchy model. Using the model to get the characteristic component of the input data, and then the new training sample set is obtained, as follows:$M'=\{(\boldsymbol{X_m},\boldsymbol{Y_m})|m=1,2,\ldots M;\boldsymbol{X_m}\in\boldsymbol{R}^{\boldsymbol{d}'_1};\boldsymbol{Y_m}\in\boldsymbol{R}^{\boldsymbol{d}_2}\}$, where the input vector of each training sample is $\boldsymbol{X}'_{\boldsymbol{m}}=[x_{m1},x_{m2},\ldots,x_{md'_1}]$,$m=1,2,\ldots M$; which contains d'_1 input properties.

Step 3: M' has been defined as the training sample set of the ELM $\boldsymbol{X}'_{\boldsymbol{m}}=[x_{m1},x_{m2},\ldots,x_{md'_1}]$,$m=1,2,\ldots M$; and $[\boldsymbol{Y_m}=[y_{m1},y_{m2},\ldots,y_{md_2}]$, $m=1,2,\ldots M$; represent the input vector and target value of the training sample, respectively. Thus, the number of nodes in the input layer and output layer can be assigned as d'_1 and d_2.

Step 4: Initialize the ELM parameters.

Step 5: According to the formula (4) to calculate the output weight matrixβ;

Step 6: Calculate the actual output of the FAHP-ELM network by using Eq. (1);

Step 7: Compare the actual output with the expected output, and calculate the relative error of the training of the FAHP-ELM

Step 8: Select a set of data that is different from the training sample data as the generalization sample set to measure the prediction accuracy of the proposed FAHP-ELM model, return step 1, and calculate the relative error of the generalization network.

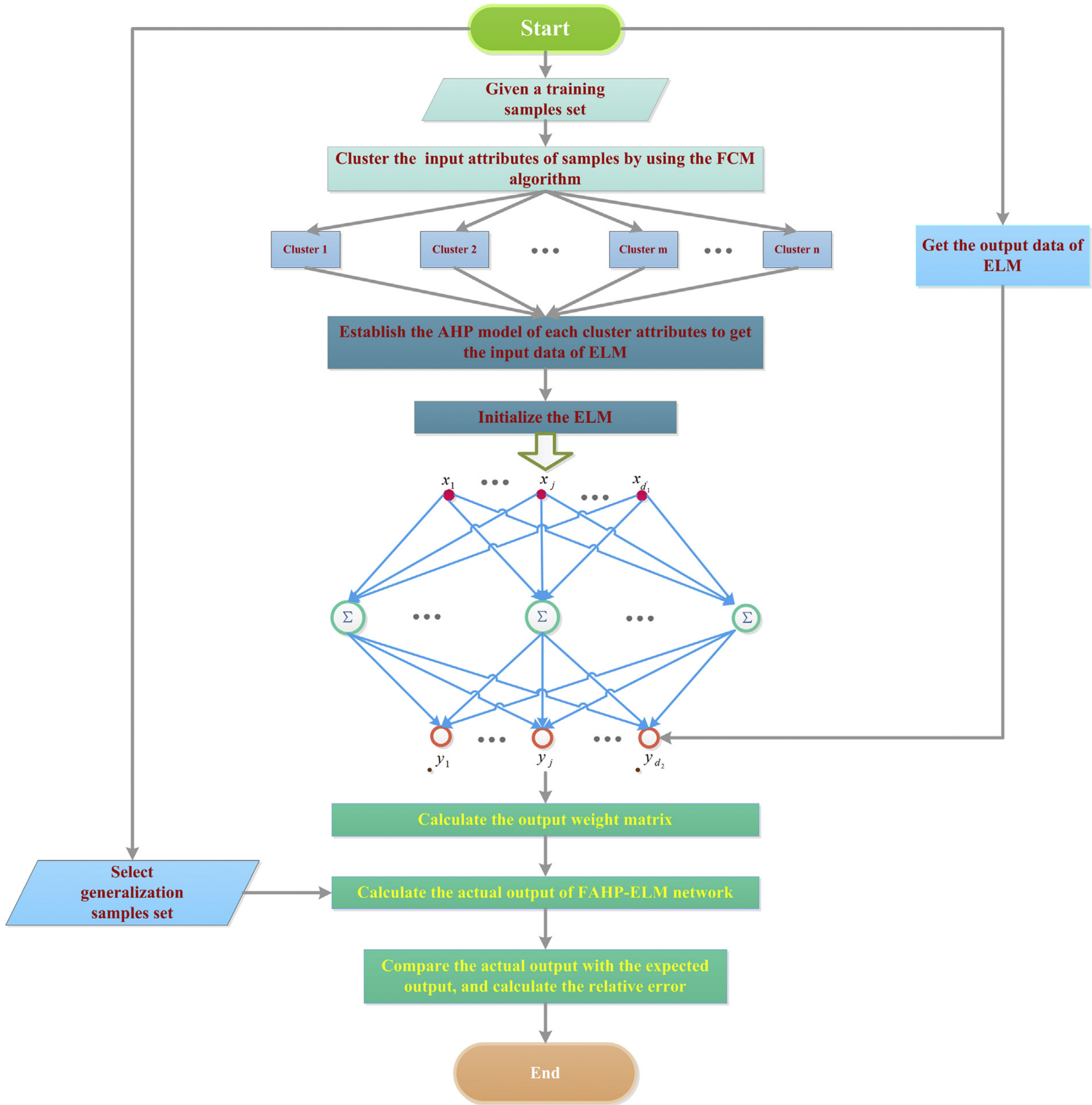

Fig. 2. Flow chart of the FAHP-ELM.

The flow chart of the FAHP-ELM is shown in Fig. 2.

4. UCI benchmark data set test

In order to verify the feasibility and validity of the proposed FAHP-ELM, the Wine Quality and the Parkinsons telemonitoring data sets in UCI database (For the details of the dataset, one can refer to the following website: http://archive.ics.uci.edu/ml/datasets.html) are selected to do the test, which is shown in Table 1.

First, cluster the input attributes of two data sets using the FCM into 5 classes and 7 classes respectively, and then do the test according to modeling process described in Section 2. Meanwhile, use the ELM and the BP neural network to model on the above experiment data, and compare the test performance of the FAHP-ELM with that of these two neural networks. We set the learning factor of the BP neural network as 0.01, the momentum factor as 0.9 and the training times as 1000. The prediction performance of

Table 1
UCI data sets.

Data Sets	Samples		Attributes	
	Training	Testing	Inputs	Output
Wine Quality	100	50	11	1
Parkinsons telemonitoring	250	91	18	1

Table 2
Test results of the WINE standard data set.

	FAHP-ELM	ELM	BP
Hidden nodes number	15	35	35
Generalization ARGE	0.061	0.097	0.083
Generalization RMSE	0.045	0.080	0.065

Table 3
Test results of the Parkinsons standard data set.

	FAHP-ELM	ELM	BP
Hidden nodes number	25	55	55
Generalization ARGE	0.052	0.063	0.059
Generalization RMSE	0.049	0.067	0.053

network is evaluated by using the average relative generalization error (ARGE) and the root mean square error (RMSE).

The test results on two kinds of UCI standard data sets of three neural networks are respectively recorded in Table 2 and Table 3.

From Tables 2 and 3, the FAHP-ELM can provide smaller values while needs less number of hidden layer nodes for getting the optimal performance. It can be seen clearly that the proposed model has higher generalization accuracy than the ELM and the BP neural network. Thus it is primarily proven of validity of the FAHP-ELM.

5. Case study: production capacity prediction and energy saving analysis of the petrochemical industry

In order to verify the actual value of the FAHP-ELM for petrochemical industries, the acetic acid consumption of the PTA solvent system and the product output of ethylene production can be predicted by the FAHP-ELM, as well as the production status and energy efficiency status of the ethylene production can be analyzed.

5.1. Predictive modeling and energy saving of the PTA solvent system

In the PTA industry, the acetic acid consumption is an important indicator to measure whether the PTA plant is advanced or not. At the same time, the consumption of acetic acid is also a main indicator to measure whether the technology is optimal or not. Therefore, reducing the consumption of tower top acetic acid in the solvent system is an important goal for production department to optimize the production of the PTA production process, improving its economic efficiency [11]. Therefore, the FAHP-ELM predicts the acetic acid consumption, analyzing the production status of the PTA plant. A schematic flow diagram of a PTA plant is shown in Fig. 3.

First, through the analysis of the operating characteristic of the solvent dehydration tower, there are 17 factors that would impact the content of acetic acid on the top tower during the PTA production process [41]. The 17 input variables are shown in Table 4. Fig. 4 described the temperature variables of the PTA plants in the actual production.

Therefore, when modeling predicting the PTA solvent system, the 17 major factors can be selected as input variables, and because of the acetic acid content in top tower is not easy to measure directly, so electrical conductivity data, which could reflect the tower top acetic acid content, is adopted as output variable of the modeling, to reflect changes of the acetic acid content indirectly, and constitute the set training data with 17 Input and 1 Output, then, based on the FCM algorithm, input data are clustered into seven categories and use the AHP algorithm based on the entropy weights to fusion the each cluster data. The fusion data is considered as the input data of the ELM, whereby the input attributes reduce from the initial seventeen to seven, and finally, establish the FAHP-ELM, analyze and forecast the production of the PTA plant and give operating instructions.

This paper makes selections of 120 groups statistical model data, and provides predictions for 60 groups of overhead acetic acid content of the PTA solvent system. At the same time, based on the above experimental data, we use the BP neural network and the ELM modeling excellently. Initialization parameter values, training and generalization precision value of three models show in Table 5.

In the process of prediction, the generalization accuracy of neural network will be changed with the increase of hidden neurons. Fig. 5 described the relationship between the error of testing and the number of hidden neurons of different networks. And the testing error of the PTA between models has been demonstrated in Table 6.

From Fig. 5 and Table 6, we can learn that the two prediction models obtained the minimum generalization error when the number of hidden neurons were 55 and 50 respectively, and the average relative error are 0.49% and 0.53% respectively. However, the generalization average relative error of the FAHP-ELM is 0.45%

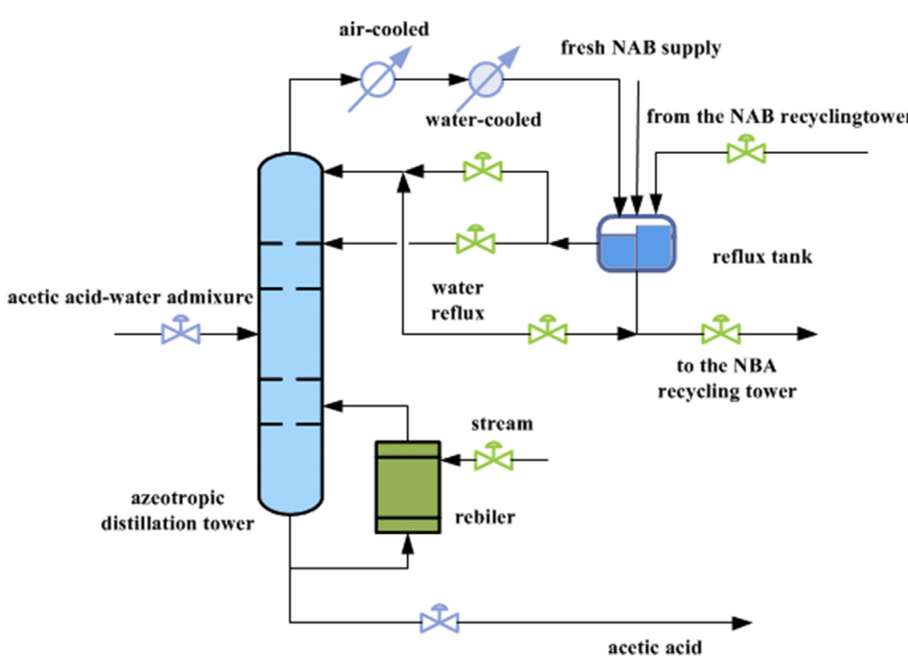

Fig. 3. Schematic flow diagram of a PTA plant.

Table 4
The input variables in the PTA plant.

No.	Input Variable
1.	feed composition
2.	feed quantity
3.	water reflux
4.	NBA main reflux
5.	NBA side reflux
6.	steam flow
7.	produced quantity of the top tower
8.	feed temperature
9.	reflux temperature
10.	temperature of the top tower
11.	temperature point above the 35th tray
12.	temperature point between the 35th tray and the 40th tray
13.	temperature point between the 44th tray and the 50th tray
14.	tray temperature near the up sensitive plate
15.	tray temperature near the low sensitive plate
16.	temperature point between the 53rd tray and the 58th tray
17.	reflux tank level

while its hidden layer has 25 neurons. It is known from the above analysis that the model based on the FAHP-ELM could reach higher generalization accuracy as well as the brief network structure together during treating with the complicated PTA solvent system data, to increase the accuracy for predicting PTA industrial capacity.

The comparison of predicted results among the ELM, the FAHP-ELM and the BP network is shown in Fig. 6.

From that, acetic acid actual contents of the first 20 sample points are commonly larger than predicted values of the FAHP-ELM. For example, the conductivity of 18th sample point actual output is 48.95 and the predicted value from model is 48.4, which means that the device at this moment operates without full load in the PTA manufacturing process and the energy efficiency level is low. But from 25th to 50th sample points, acetic acid actual contents are close to even lower than predicted values from model, which means that the device at this moment operates in the more perfected production status and energy efficiency level improves. It is known from the above analysis that the production configurations of devices for the last 25 sample points are more reasonable. These 25 groups of production parameters should be used for reference in the future production to make chemical industry production more efficiency and increase energy utilization. For example, in order to realize the energy savings, the changes of production parameters of 18th sample point are shown in Fig. 7.

As shown in Fig. 7, in order to make production more effectively and increase energy utilization, the parameters No.10, No.11, No.12,

Table 5
Setting the initial value of models parameters.

Model parameters	BP	ELM	FAHP-ELM
Input layer node number	17	17	7
Output layer node number	1	1	1
Learning factor	0.01	–	–
Momentum factor	0.7	–	–
Maximum iterations	1000	–	–
Activation function	Sigmoid	–	–

No.13 and No. 15 should be reduced 0.314, 0.0091, 0.283, 0.998 and 0.000658, respectively. Meanwhile, the parameters of No.14 and No.16 should be increased input 0.005 and 0.0736, then according to the above analysis to adjust production parameters, the plant can reduce the consumption of acetic acid 0.588.

5.2. Predictive modeling and efficiency analysis for the ethylene industrial production

Monthly production data from 2010 to 2014 in 19 ethylene production plants of the whole country with main 7 kinds of ethylene production technologies are selected as analyzing objects [42,43].

First, we would introduce the ethylene industry production briefly. The ethylene industry is a symbol of the process industry. Its energy consumption accounted for more than half of the process industrial consumption. And ethylene production plants are a key member of fossil fuel consumed energy sources that can be divided into two parts: cracking and separation, where the production and operation of cracking furnace need to consume a large amount of fuel. And the separation section mainly contains three parts: a rapid cooling, a compression and a separation part [44]. In ethylene production, the different ethylene energy efficiency analysis battery limits and computational methods are used by the different enterprises. Therefore, this article references ethylene trade standard DB 37/751–2007 and GB/T 2589-2008 [45,46]to perform the division of ethylene production plant battery limit. A schematic flow diagram of an ethylene plant is shown in Fig. 8.

As for ethylene plants, the factors which directly related to the production efficiency mainly include: (1) raw materials; (2) fuel and consumption of power; (3) products. It is known from the energy battery limit of ethylene production that: the crude oils including naphtha, light diesel fuel, raffinate, hydrogenation tail oil, light hydrogenation tail oil, C_{345} and others; fuel including light weight oil, heavy weight oil and fuel gas; steams including super

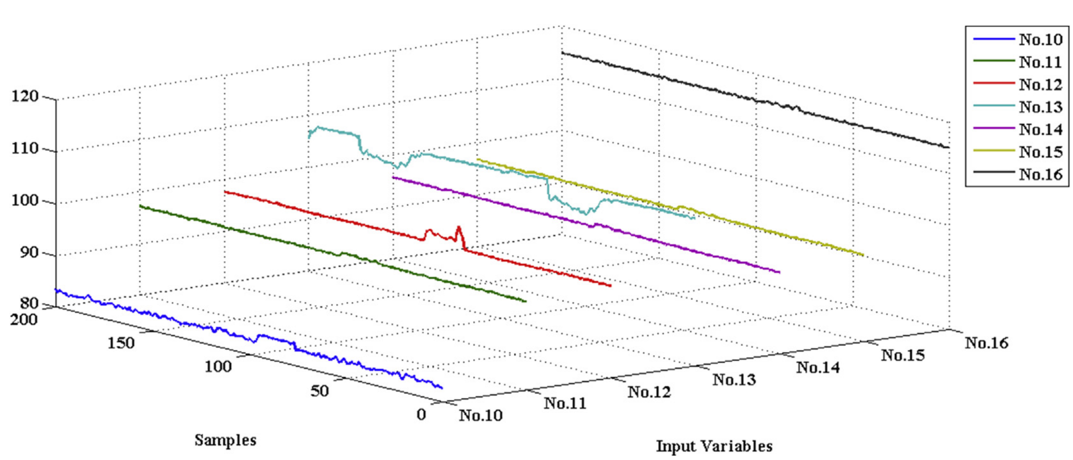

Fig. 4. The temperature variables of PTA plants in the actual production.

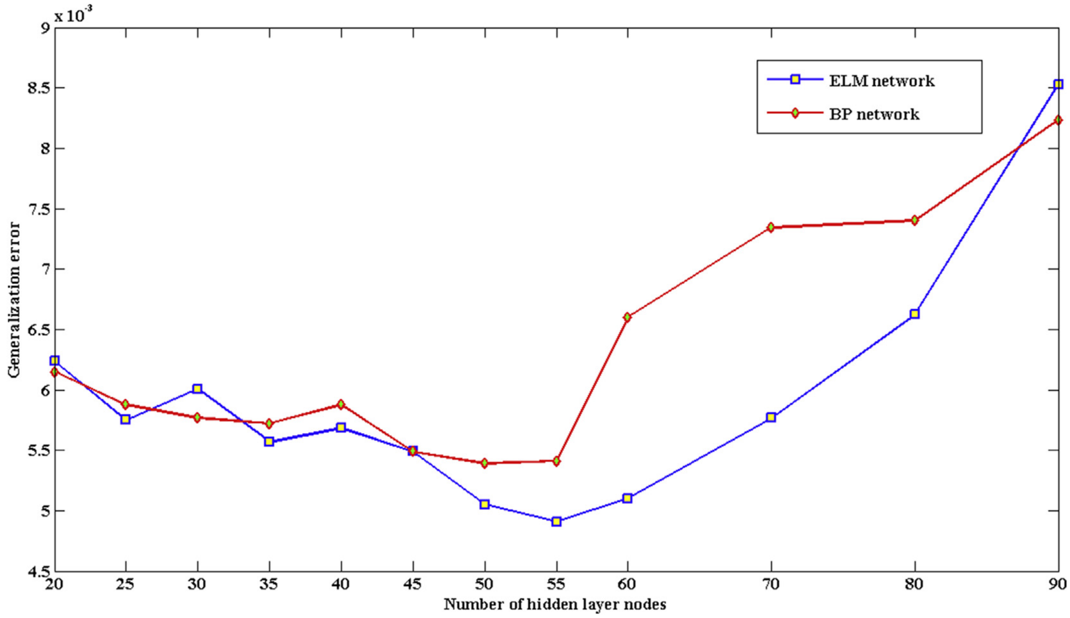

Fig. 5. Relation chart between error of testing and number of hidden neurons.

Table 6
Testing error of the PTA between models.

		FAHP-ELM	ELM	BP
	Hidden nodes number	25	55	50
Training	Average relative error/%	0.43	0.34	0.44
	Training times/s	0.13	3.58	22
Testing	Average relative error/%	0.45	0.49	0.53

high pressure steam, high pressure steam, medium pressure steam and low pressure steam; water including circulating water, industrial water, boiler water and other water and the electricity. And the consumption of crude oils, fuel, water, steams and electricity are taken as the inputs of ethylene production, and the sum of main products as ethylene, propylene and C_4 are the output indexes [32,42]. Meanwhile according to conversion relations of Table 3.0.2 and Table 3.0.3 in Computational Method of Petrochemical Design Energy Consumption (SH/T3110-2001) [47], the measure unit of energy consumption parameters of steams, electricity, water and fuel are translated into the GJ. And the production of crude oil, ethylene, propylene and C4 are measured in tons. Figs. 9 and 10 described the required fuels, steams, water, electricity and crude oil and output of some partial ethylene plants.

We selected 200 sets of actual statistical data in 2010 for modeling and predicted 82 sets of ethylene plant output in 2011. And according to the analysis of ethylene product system, the number of clustering categories in the FAHP algorithm will be set to five. At the same time the above experiment data is modeled by the

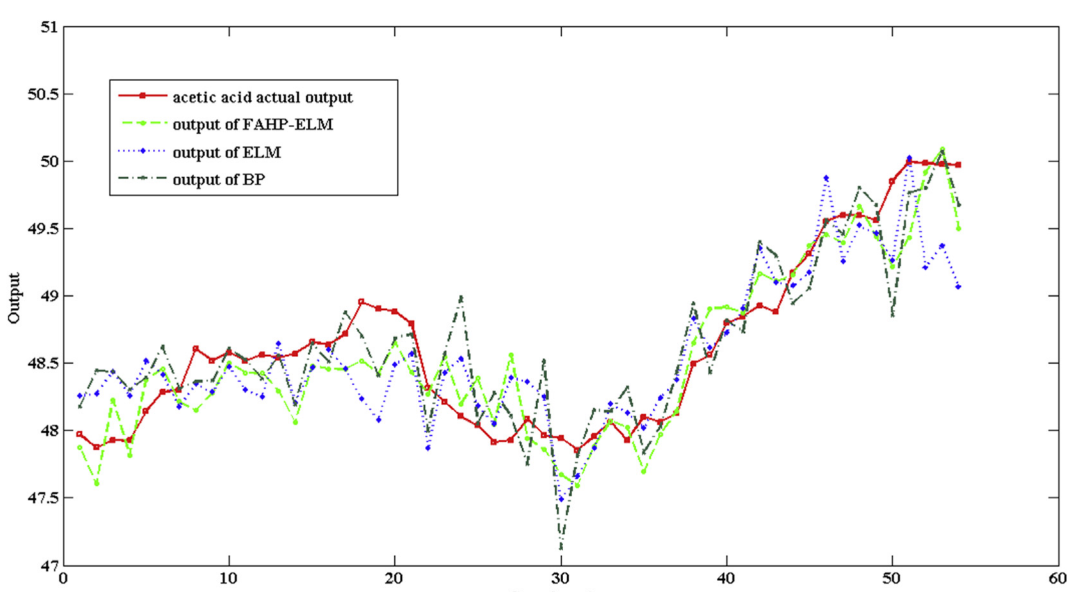

Fig. 6. Comparison of predicted results among the ELM, the FAHP-ELM and the BP.

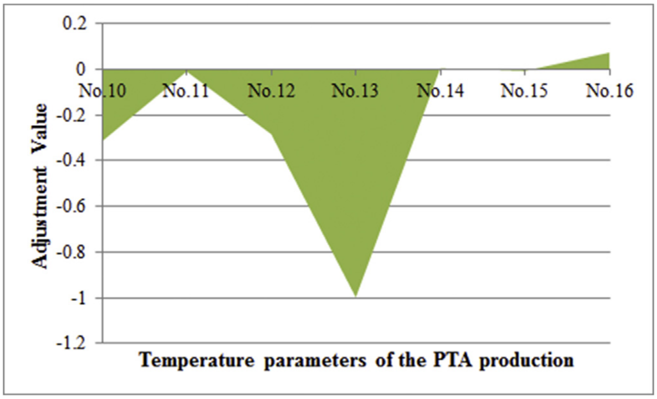

Fig. 7. The adjustment of production parameters of 18th sample point.

BP neural network and the ELM to compare the forecasting performance with the FAHP-ELM. So the initial values of three kinds of network parameters are shown respectively in Table 7.

Fig. 11 described the relationship between the error of testing and the number of hidden neurons of different networks. And the testing error of ethylene production between networks has been demonstrated in Table 8.

It can be seen from the result in Fig. 11 and Table 8 that the two prediction models yields the minimum generalization error when the number of hidden neurons is 22 and 30 respectively, and the average relative error is 5.26% and 6.99% respectively. However, the generalization average relative error of the FAHP-ELM is 2.68% while its hidden layer has 12 neurons. Meanwhile, as for the required time of training process of 3 networks, the FAHP-ELM is obviously faster, the training time of which is 0.31% of the BP neural network and 54.38% of the ELM. And Fig. 12 demonstrates the comparison chart of predicted results among the ELM, the FAHP-ELM and the BP network.

It is known from the above analysis that the predicted model based on FAHP-ELM could reach better performance of generalization as well as the brief network structure together during treating with the complicated ethylene industrial production data, to increase the accuracy for predicting ethylene industrial capacity.

Based on the energy efficiency data in 2012 from two 800 thousand tons ethylene plants with two different kinds of technology, this paper uses the data to build the predictive model and predict the total output of ethylene, propylene and C4 in 2013. Table 9 and Fig. 13 show the predicted result and the comparison of generalization errors of two plants.

Table 9 and Fig. 13 show that the error of Plant 2 is less than that of Plant 1, and this result shows that the production condition of Plant 2 is good and stable which is much better than that of Plant 1. Based on the data and the condition, Plant 1 should adjust the input of ethylene and use the production technology of Plant 2 to improve the energy efficiency of the ethylene production. For example, Plant 1 uses 238,440-ton crude, 19.23GJ fuel, 0.49GJ stream, 2.10GJ water and 0.6GJ electricity to produce ethylene, propylene and C4 in May, 2013. The actual value of the total output is 106,212 tons, and the predicted value is 118,009 tons. Based on the data above, Plant 1 has low

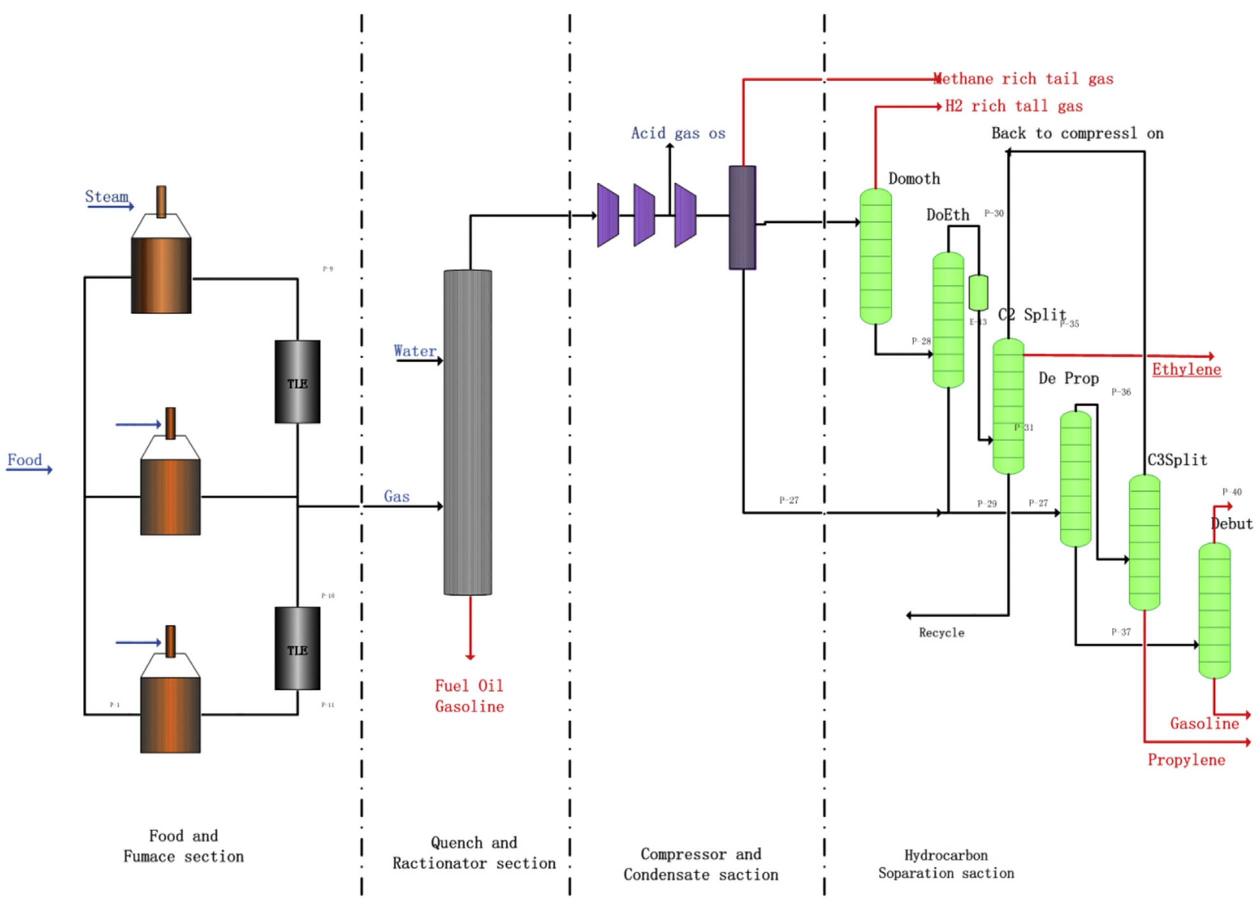

Fig. 8. Schematic flow diagram of an ethylene plant.

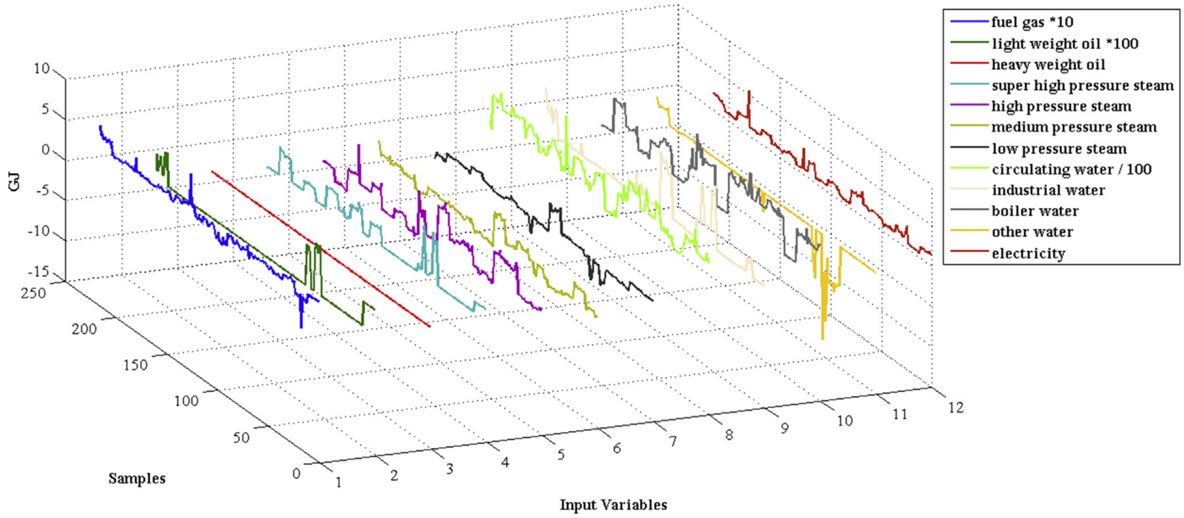

Fig. 9. Energy types (water, steams, fuels, electricity) of ethylene plants in 2010.

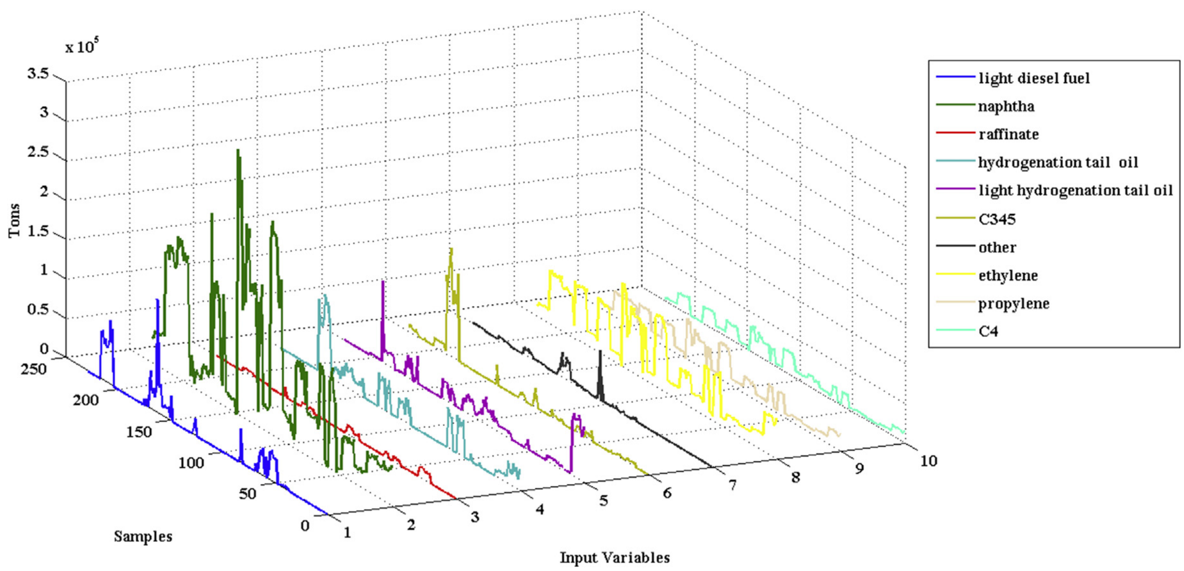

Fig. 10. The required crudes and the production of ethylene, propylene and C4 of ethylene plants in 2010.

energy efficiency in May and should adjust the input to improve the efficiency of production in the future. Meanwhile, Plant 2 uses 185,995-ton crude, 18.81GJ fuel, 1.29GJ stream, 2.69GJ water and 1.93GJ electricity to produce ethylene, propylene and C4. The actual value of total output is 96,851 tons, and the predicted value is 98,605 tons. Based on the data above, the production condition is good in June and should be kept in the future. Meanwhile, the production conditions of plant 1 should be adjusted, so the adjust of production parameters of Plant 1 are shown in Fig. 14 and Fig. 15.

Table 7
Initial value setup of models parameters.

Model parameters	BP	ELM	FAHP-ELM
Node No. of input layer	19	19	5
Node No. of output layer	1	1	1
Learning factor	0.01	–	–
Momentum factor	0.7	–	–
Max. iteration	1000	–	–
Activation function	Sigmoid	–	–

From Figs. 14 and 15 we can learn that the Plant 1 can enhance the efficiency of energy utilization to achieve energy conservation and emission reduction by increasing the input of stream, water and electricity 0.8GJ, 0.59GJ and 1.33GJ respectively. At the same time, the input of crude oil and fuel should be reduced to 52445 tons and 0.42 GJ. On the other hand, Plant 1 need 2.24-ton crude and Plant 2 needs 1.92-tons crude oil to manufacture each ton of production. According to carbon emission factor for different types of fuels [48], comparing with Plant 2, Plant 1 will produce 0.546-tons more CO_2 emission at the same time. The CO_2 emission can be reduced by 436,800 tons when the scale of production is 800000 tons, on the premise that the production department adjust production parameters of Plant 1 according to the above analysis.

6. Conclusions

The paper proposes the FAHP-ELM prediction model. This

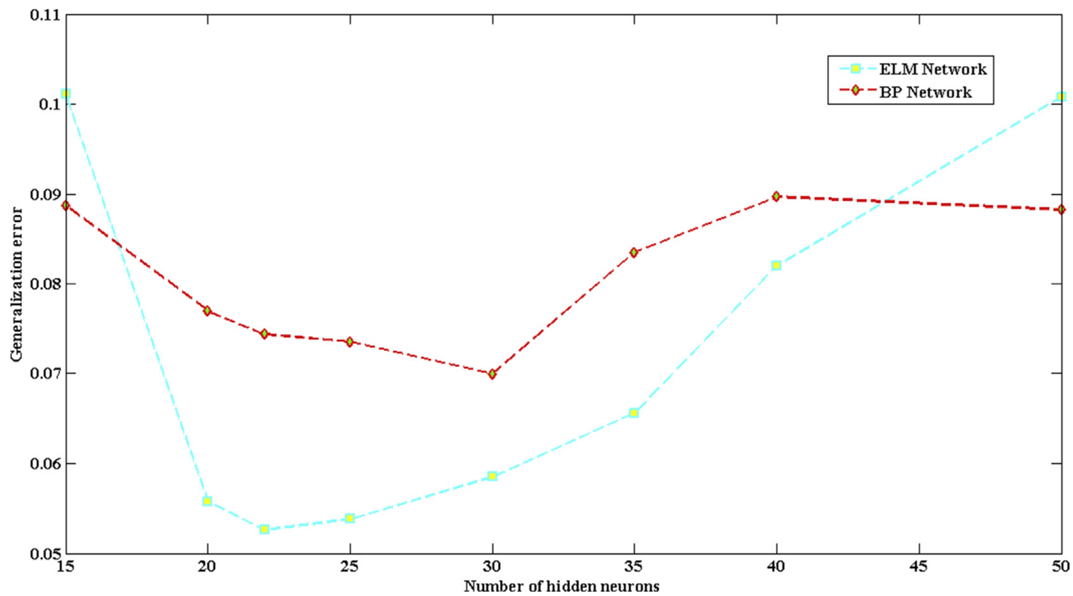

Fig. 11. Relation chart between the error of testing and the number of hidden neurons.

Table 8
Testing error of the ethylene industrial between models.

		FAHP-ELM	ELM	BP
Training	Hidden nodes number	12	22	30
	Average relative error/%	2.42	5.01	4.28
	Training times/s	0.0062	0.0114	2
Testing	Average relative error/%	2.68	5.26	6.99

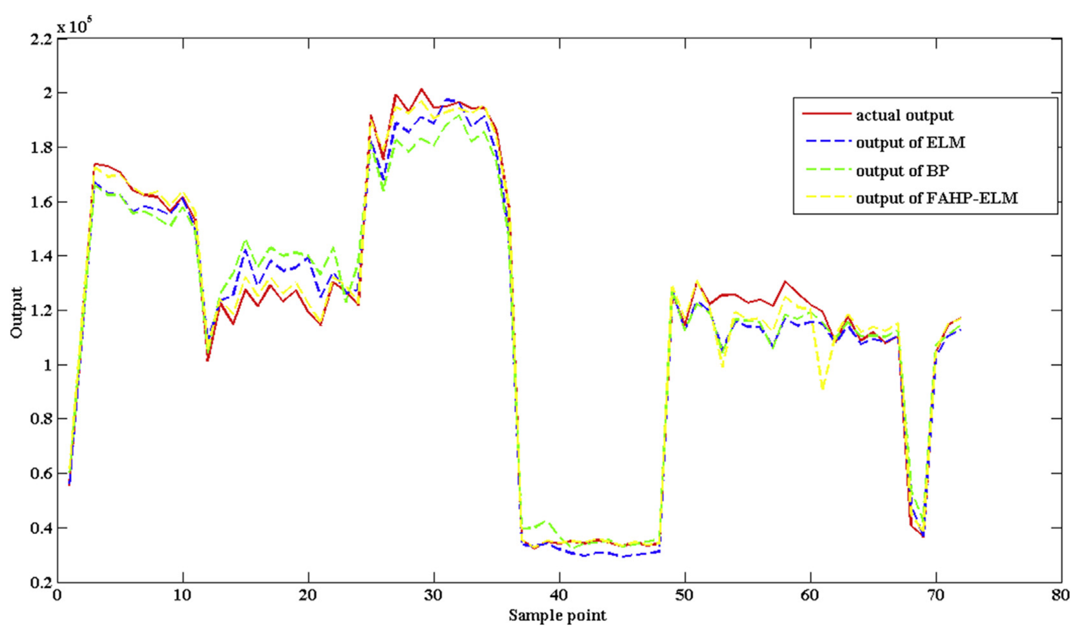

Fig. 12. Result comparison of the ethylene industry.

proposed model is able to filter redundant information and extract characteristic components. Meanwhile, the FAHP-ELM is also capable of training the blended data, simplifying the network structure of the ELM as well as optimizing the learning performance of networks. Compared with the BP neural network and the

Table 9
The predicted relative error of two ethylene plants.

	Plant 1	Plant 2
Predicted relative error/%	4.9	1.8

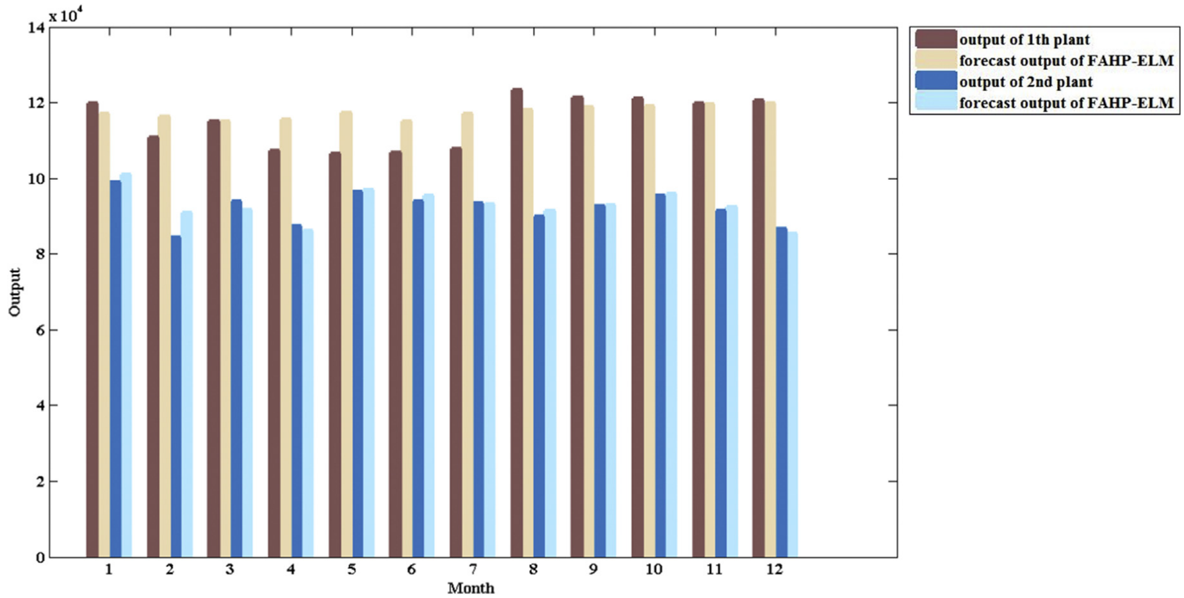

Fig. 13. The predicted result of two ethylene plants.

ELM, the faster convergence, higher effectiveness and accuracy, and stronger generalization of modeling of the FAHP-ELM are verified by UCI standard data sets. Meanwhile, the proposed model is an application in the forecast modeling analysis of petrifaction production data, it can accurately predict the production of petrifaction production, objectively analyze the energy status in petroleum chemical plants. Moreover, the experimental results demonstrate that the FAHP-ELM could reduce the consumption of acetic acid by guiding the production department more reasonably to configuration parameters in PTA production. And then the FAHP-ELM, reasonably allocates input of crude oils, streams, fuel, water and electricity in order to improve the production efficiency, save energy and reduce the carbon emissions. Furthermore, the FAHP-ELM is the theoretical basis and has more practical value for other complex industrial prediction modeling.

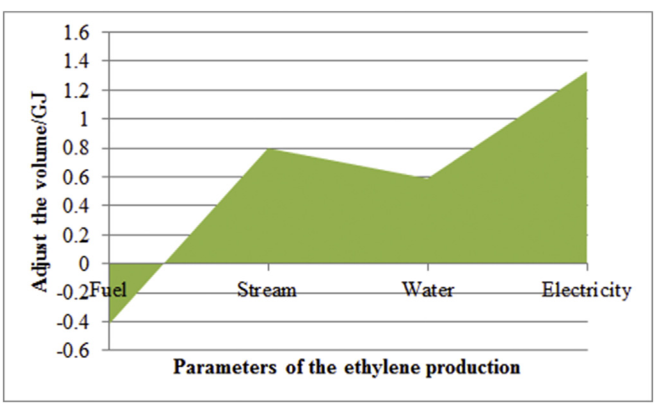

Fig. 14. The adjustment of production parameters (water, steams, fuels, electricity) of ethylene Plant 1.

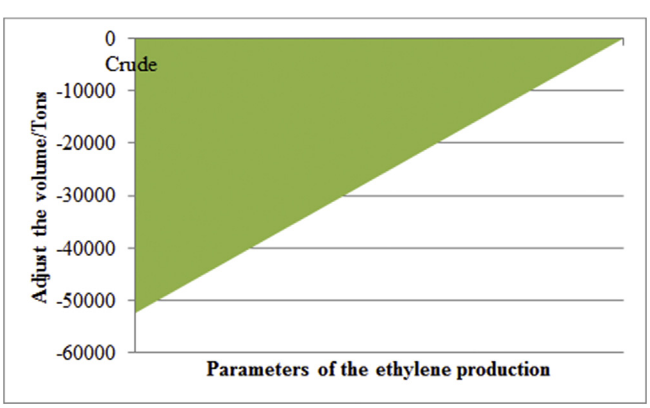

Fig. 15. The adjustment of production parameters (crude oil) of ethylene Plant 1.

However, the FAHP-ELM is applied to analyze and predict the energy status of petrochemical plants effectively, and the parameters of model are adjusted by experiments. Therefore, we will improve our model that parameters can be adjusted automatically, such as designing a self-organizing FAHP-ELM, which is more suitable to the real world applications.

Acknowledgments

This research was partly funded by National Natural Science Foundation of China (61533003, 61603025), Natural Science Foundation of Beijing, China (4162045) and National Key Technology Support Program (2015BAK36B04).

References

[1] Ma GF, Xu YH, Guo X. Ethylene business review of China petrochemical in 2014. Ethyl Ind 2015;27(1):1–5.
[2] Li YL, Sun XT. Development status and trend of China PTA industry. China Pet Chem Ind Anal 2013;(8):46–9.
[3] Qian BZ. Technical progress and market analysis of PTA. Polyest Ind 2012;25(1):11–6.
[4] Muyeen SM, Hasanien HM, Al-Durra A. Transient stability enhancement of wind farms connected to a multi-machine power system by using an adaptive ANN-controlled SMES. Energy Convers Manag 2014;78:412–20.
[5] Wang JZ, Jiang H, Wu YJ, Dong Y. Forecasting solar radiation using an optimized hybrid model by Cuckoo Search algorithm. Energy 2015;81:685–95.
[6] Sun W, Xu YF. Financial security evaluation of the electric power industry in China based on a back propagation neural network optimized by genetic algorithm. Energy 2016;101:366–79.
[7] Aghbashlo M, Shamshirband S, Tabatabaei M, Yee PL, Larimi YN. The use of ELM-WT (extreme learning machine with wavelet transform algorithm) to predict exergetic performance of a DI diesel engine running on diesel/

biodiesel blends containing polymer waste. Energy 2016;94:443–56.

[8] Mohammadi K, Shamshirband S, Yee L, Petković D, Zamani M, Ch S. Predicting the wind power density based upon extreme learning machine. Energy 2015;86:232–9.

[9] Monedero I, Biscarri F, León C, Guerrero JI, González R, Pérez-Lombard L. Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant. Expert Syst Appl 2012;39(10):9860–7.

[10] Avramović JM, Veličković AV, Stamenković OS, Rajković KM, Milić PS, Veljković VB. Optimization of sunflower oil ethanolysis catalyzed by calcium oxide: RSM versus ANN-GA. Energy Convers Manag 2015;105:1149–56.

[11] He YL, Geng ZQ, Zhu QX. Data driven soft sensor development for complex chemical processes using extreme learning machine. Chem Eng Res Des 2015;102:1–11.

[12] Haghbakhsh R, Adib H, Keshavarz P, Koolivand M, Keshtkari S. Development of an artificial neural network model for the prediction of hydrocarbon density at high-pressure, high-temperature conditions. Thermochim Acta 2013;551:124–30.

[13] He YL, Xu Y, Geng ZQ, Zhu QX. Soft sensor of chemical processes with large numbers of input parameters using auto-associative hierarchical neural network. Chin J Chem Eng 2015;23(1):138–45.

[14] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–6.

[15] Xie TT, Yu H, Wilamowski BM. Comparison between traditional neural networks and radial basis function networks. In: Proceedings of the IEEE International symposium on industrial electronics; 2011. p. 1194–9.

[16] Yu H, Xie TT, Paszczynski S, Wilamowski BM. Advantages of radial basis function networks for dynamic system Design. IEEE Trans Ind Electron 2011;58(12):5438–50.

[17] Xu Y, Feng J, Zhu QX. Research and application of extension theory-based radial basis function neural network. Control Decis 2011;26(11):1721–5.

[18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE Int Jt Conf IEEE 2004;2:985–90.

[19] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1):489–501.

[20] Miche Y, Akusok A, Veganzones D, Björk KM, Séverin E, Jardin P, et al. SOM-ELM—self-organized clustering using ELM. Neurocomputing 2015;165: 238–54.

[21] Huang GB, Zhou HM, Ding XJ, Zhang R. Extreme learning machine for regression and multiclass classification. Syst Man Cybern Part B Cybern IEEE Trans 2012;42(2):513–29.

[22] Sun ZL, Wang H, Lau WS, Seet G, Wang DW. Application of BW-ELM model on traffic sign recognition. Neurocomputing 2014;128:153–9.

[23] Zhang Y, Zhang L, Li P. A novel biologically inspired ELM-based network for image recognition. Neurocomputing 2016;174:286–98.

[24] Tang J, Deng C, Huang GB. Extreme learning machine for multilayer perceptron. IEEE Trans Neural Netw Learn Syst 2016;27:809–21.

[25] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—Extreme learning machine approach. Energy Convers Manag 2014;87:10–8.

[26] Cao JW, Lin ZP, Huang GB. Self-adaptive evolutionary extreme learning machine. Neural Process Lett 2012;36(3):285–305.

[27] Lan Y, Soh YC, Huang GB. A constructive enhancement for online sequential extreme learning machine. International Joint Conference on Neural Networks. IEEE Press; 2009. p. 1708–13.

[28] Tang JX, Deng CW, Huang GB. Extreme learning machine for multilayer perceptron. IEEE Trans Neural Netw Learn Syst 2015;27(4):1.

[29] Gastaldo P, Bisio F, Decherchi S, Zunino R. SIM-ELM: connecting the ELM model with similarity-function learning. Neural Netw 2016;74:22–34.

[30] Han YM, Geng ZQ, Zhu QX, Lin XY. Energy consumption hierarchical analysis based on interpretative structural model for ethylene production. Chin J Chem Eng 2015;23(12):2029–36.

[31] Geng ZQ, Wang Z, Peng CL, Han YM. A new fuzzy process capability estimation method based on kernel function and FAHP. IEEE Trans Eng Manag 2016;63(2):177–88.

[32] Han YM, Geng ZQ, Liu QY. Energy efficiency evaluation based on data envelopment analysis integrated analytic hierarchy process in ethylene production. Chin J Chem Eng 2014;22(12):1279–84.

[33] Yagmur L. Multi-criteria evaluation and priority analysis for localization equipment in a thermal power plant using the AHP. Energy 2016;94:476–82.

[34] Pham DL. Spatial models for fuzzy clustering. Comput Vis Image Underst 2001;84(2):285–97.

[35] Oliveira JVD, Pedrycz W. Advances in fuzzy clustering and its applications. West Sussex: Wiley; 2007.

[36] Pimentel BA, Souza RMCRD. Multivariate Fuzzy C-Means algorithms with weighting. Neurocomputing 2016;174:946–65.

[37] Zainuddin Z, Pauline O. An effective fuzzy C-means algorithm based on symmetry similarity approach. Appl Soft Comput 2015;35:433–48.

[38] Zhang LY, Lu W, Liu XD, Pedrycz W, Zhong CQ. Fuzzy C-Means clustering of incomplete data based on probabilistic information granules of missing values. Knowledge Based Syst 2016;99:51–70.

[39] Pal NR, Bezdek JC. On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 1995;3(3):370–9.

[40] Gu XB, Zhu QX. Difference driving fuzzy decision algorithm based on fuzzy eigen vector. Comput Eng Appl 2005;41(2):53–6.

[41] He YL, Geng ZQ, Zhu QX. Positive and negative correlation input attributes oriented subnets based double parallel extreme learning machine (PNIAOS-DPELM) and its application to monitoring chemical processes in steady state. Neurocomputing 2015;165:171–81.

[42] Geng ZQ, Han YM, Gu XB, Zhu QX. Energy efficiency estimation based on data fusion strategy: case study of ethylene product industry. Ind Eng Chem Res 2012;51:8526–34.

[43] Han YM, Geng ZQ, Gu XB, Wang Z. Performance analysis of China ethylene plants by measuring Malmquist production efficiency based on an improved data envelopment analysis cross-model. Ind Eng Chem Res 2015;54:272–84.

[44] Han YM, Geng ZQ, Zhu QX. Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry. Energy 2015;83:685–95.

[45] China standards: the general computing guide of special energy consumption (GB/T2589-2008). 2008.

[46] China standards: the limitation of energy consumption for ethylene product (DB37/751-2007). 2008.

[47] Calculation method for energy consumption in petrochemical engineering Design (SH/T3110-2001). 2002.

[48] Zhao M, Zhang WG, Yu LZ. Carbon emissions from energy consumption in Shanghai city. Res Environ Sci 2009;22(8):984–9.