Energy 160 (2018) 898–909

Contents lists available at ScienceDirect

Energy

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

Production prediction and energy-saving model based on Extreme Learning Machine integrated ISM-AHP: Application in complex chemical processes

Zhiqiang Geng [a, b], Hongda Li [a, b], Qunxiong Zhu [a, b], Yongming Han [a, b, *]

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

ARTICLE INFO

Article history:
Received 9 March 2018
Received in revised form
12 July 2018
Accepted 13 July 2018
Available online 17 July 2018

Keywords:
Interpretative structural modeling (ISM)
Analytic hierarchy process (AHP)
Extreme learning machine (ELM)
Production capacity prediction
Energy-saving
Complex chemical processes

ABSTRACT

One of the key issues of the sustainable development in all countries is industrial productivity improvement, especially the production capacity improvement and energy-saving of complex chemical processes. Therefore, this paper proposes a production prediction and energy-saving model based on Extreme Learning Machine (ELM) integrated Interpretative Structural Modeling (ISM) and Analytic Hierarchy Process (AHP). The factors that affect the productivity are divided into different levers by the ISM. And then the attributes of each layer are fused by the AHP based on the entropy weight, which greatly reduces the complexity of the input attributes. Moreover, the production prediction and energy-saving model is established based on the ELM. Compared with the traditional ELM, the validity and the practicability of the proposed method are verified by University of California Irvine (UCI) datasets. Finally, the proposed method is applied in the production capacity prediction and energy-saving of ethylene production systems and Purified Terephthalic Acid (PTA) production systems. The experimental results show the proposed method could reduce the number of hidden layer nodes and improve the training time of the ELM. Furthermore, the prediction accuracy of the ethylene production and the PTA production reaches about 99% to improve the energy efficiency of complex chemical processes.

1. Introduction

The sustainable development strategy has played an important guiding role in building a well-off society in an all-round way. However, China is a country with a large population, the shortage of natural resources and a relatively backward condition in economy and technology. Therefore, the social and economic virtuous circle can be achieved through the resource rational utilization and the environmental protection, especially complex chemical processes, which can promote the unity of ecological, economic and social benefits. Ethylene production capacity is an important indicator reflecting the development of a national petrochemical industry [1]. As is reported by China Petroleum and Natural Gas Co, Ltd. in 2015 [2], the annual production of ethylene was 11005.2 kt (not including the Central Plains Petrochemical MTO device 112.7 kt), which increased 418.9 kt compared with the same period last year. And the ethylene energy consumption (standard oil) was 559.06 kg/t, which decreased 10.01 kg/t. The "12th Five-Year Plan" has reduced energy consumption to less than 600 Tons/Ton ethylene. However, the domestic high energy consumption of ethylene products caused by the integrated cost is too high. Energy consumption in the comprehensive cost occupies a considerable proportion. According to the 2016–2021 China ethylene production equipment industry market demand and investment advisory report, the energy consumption of the ethylene device topped the list [3]. Meanwhile, compared with foreign countries, there are still some gaps in energy-saving [4].

In addition, Purified terephthalic acid (PTA) is the bottom product of the oil industry, and 90% of the PTA around the world is used to produce polyester. Polyester is the important basic chemical raw material of China, which has a large market demanding. In recent years, the production capacity of PTA in China is constantly expanding [5]. However, there is still great potential for the improvement in energy savings and emission reduction [6]. Therefore, production prediction and energy-saving of the ethylene

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

https://doi.org/10.1016/j.energy.2018.07.077
0360-5442/

production and the PTA production in complex chemical processes is of great significance to the conservation of oil resources and the sustainable development strategy.

Nowadays, many related prediction methods have been proposed. The fuzzy C-means algorithm is used to cluster high-dimensional data, but the quality of clustering is affected by the initial value setting, the number of clusters and the selection of parameters [7]. The curve fitting model also has a relatively positive effect in fossil fuel production prediction, but the constraints are also more complex [8]. The crude oil production is predicted by using the artificial neural network (ANN) prediction model [9], but the prediction error is high. Moreover, the training efficiency of the ANN prediction model is lower for the multi-attribute large data.

Therefore, this paper proposes a production prediction and energy-saving model based on Extreme Learning Machine (ELM) integrated the interpretative structural modeling method (ISM) and the analytic hierarchy process (AHP). First, the input attributes are classified into different levers by the ISM based on partial correlation coefficient. Second, the input attributes of each layer are merged by using the AHP based on the entropy weight to reduce the number of input attributes of the ELM. Moreover, the production capacity prediction and energy-saving model is obtained by the ELM. Finally, the validity and practicability of the proposed method is validated by University of California Irvine (UCI) datasets. Furthermore, the proposed method is applied to the production prediction and energy-saving of the ethylene production system and the PTA production system in complex chemical processes. The experimental results show that the proposed method can reduce the number of hidden nodes and improve the training efficiency. Meanwhile, the prediction accuracy of the ethylene production and the PTA production reaches about 99% to improve the energy efficiency.

The remaining parts are organized as follows: Section 2 introduces the research status of production prediction and energy saving. In Section 3, the improved method is described in detail. Compared with other feedforward neural networks, the validity and practicability of the proposed method is validated by UCI datasets in Section 4. Section 5 describes the energy efficiency analysis and two prediction examples of complex chemical industries based on the improved method. Finally, the discussion and the conclusion are obtained in Section 6 and in Section 7, respectively.

2. Related work

Nowadays, the development of the ANN brings the gospel to many complex problems [10]. And the potential and non-linear relationship of the complex problem can be neglected based on the ANN. Meanwhile, the potential relationship between the input and the output in a certain error range makes a quick identification. Although the ANN has some requirements in the quality and quantity of the training data, there are common shortcomings of slow convergence, large amount of computation, long training time and unexplained. With the development of the ANN, the models have been gradually optimized [11]. Especially in the past several decades, the ANN had successfully solved many practical problems in many fields such as predicting stock market indices [12], predicting the dynamic viscosity of oxide nanoparticles suspension in water and ethylene glycol [13], developing the mobile dose prediction system [14], and energy-saving and management of complex chemical processes [15]. At present, dozens of neural network algorithms and mathematical models have been produced, which have played a key role in the development of artificial intelligence. The back propagation (BP) neural network was proposed by Rumelhart et al. [16] in 1986 to solve the problem of exclusive or (XOR). Gradient search techniques are used to minimize the mean square error (Mse) between the actual output value and the expected output value. In order to accurately describe the state of the power battery, the BP neural network with S-shaped excitation function is used to improve the accuracy of the charging state by integrating the terminal voltage deviation as feedforward and feedback [17].

Radical Basis Function (RBF) neural network has the characteristic of simple and quick features. In 1971, Hardy used the RBF to analyze topography and other irregular surfaces and get a breakthrough [18]. By optimizing the number of hidden neurons and selecting centers among a wide range of data, the RBF neural network can be trained to deliver maximum fitting and computational performance [19]. The RBF is constructed to substitute the complex forward calculations. And an adaptive approach based on cross-entropy is applied to efficiently improve the accuracy of the surrogate model [20]. In addition, the RBF has strong adaptability and positive effect on eliminating modeling errors [21].

With the continuous expansion of data, the learning speed of the BP and the RBF cannot meet the need of production industries. Therefore, to increase the learning efficiency of neural networks, the ELM was put up by Huang et al. in 2004 [22]. The ELM is a single hidden layer feedforward neural network (SLFN) whose learning rate and generalization ability is better than simple feedforward neural networks. Meanwhile, the least training errors and norm of weights can be obtained by the ELM [23]. Li et al. extracted the information from the market news and stock prices to predicate the profitability of the stock market using the limit learning machine [24]. Compared with the RBF, the ELM has a significant improvement in predicating accuracy and predicating speed. However, the ELM needs higher on the CPU resource. Nowadays, the ELM has been optimized and applied in energy prediction. The kernel-based ELM was applied in energy-saving and emission reduction engine modeling and framework optimization in biodiesel [25]. Although the data definition is relatively scarce, the ELM still shows a strong ability to predict and adapt. Therefore, this paper proposes an improved ELM based on the main factors analysis.

The ISM based on the partial correlation coefficient is a kind of structured modeling technology established by Warfield [26] from 1971 to 1973. The ISM is widely used in the alarm root cause and the abnormality propagation path [27], the supply chain resilience system model [28], and the influence factors analysis of college students' employment [29]. The use of the relationship between the graph and the matrix has solved a variety of complex problems at this stage. Zhang et al. [30] used the ISM to judge the factors that affect the network reconstruction, and solved the problem of restoring a power system after a complete blackout or a local outage. Bin et al. [31] used the structural model integrated the fuzzy analysis network process to predict the development obstacles in shale gas industry, and solve the development obstacles and strategic measures between the causal relationships. Zeng et al. [32] decomposed the structural model of the index system to identify three indicators that affect the low carbon efficiency. However, factors in different levers have different roles in energy-saving and production capacity. Geng et al. proposed the ELM integrated the ISM model to guide the improvement direction of energy-saving and carbon emissions reduction in complex petrochemical systems [10]. However, the ELM integrated the ISM model only considers the impacts of the first level of elements obtained based the ISM on the prediction result, and does not consider the mutual influence between all layers or between the same layers. In order to solve this problem, the AHP algorithm based on entropy weight is introduced. Thomas et al. [33] introduced the concept of the AHP and the consistency of the matrix data in 1977 by defining and measuring the consistency of the matrix data and the main feature

vector of the pairwise comparison matrix. In order to turn complex problem hierarchical structure into multi-standard decision-making, the AHP effectively makes large-scale problems broken down to get the maximum measurement. Meanwhile, the AHP is applied to the planning of renewable energy, and provide important information for energy investment and central energy policy formulation [34]. In order to determine the optimal placement of solar energy collection equipment, the AHP is used to arrange the factors affecting the efficiency of solar energy collection [35]. A new decentralized scheduling strategy based on AHP for electric vehicle users and power grids is applied to adjust a variety of different standards with balancing the contradictory relationship between costs of user and battery charging states [36]. Moreover, the AHP has succeeded in making contributions in large complex issues of decision-making stratification, such as the quantitative and comprehensive evaluation [37], the multi-criteria decision making techniques [38] and the compressor selection in petrochemical industry [39]. Therefore, this paper proposes an improved ELM integrated the ISM and the AHP(ISM-AHP).

3. The ELM integrated ISM-AHP

The ISM based on partial correlation coefficient is used to analyze and judge the correlation among attributes of the data to get the main factor. And then the attributes of each layer are fusion by the AHP to reduce the dimension as the input of the ELM and improve the generalization ability.

3.1. Factor analysis based on ISM-AHP

3.1.1. The ISM based on the partial correlation coefficient

The ISM analyzes the relationship between the various factors with a certain degree of information processing. The use of directed graphs and matrices on the relationship between the various factors clears the overall structure of the complex problem [27] [40].

A directed graph model is created to construct the adjacency matrix and the reachable matrix. The relationships between the elements are expressed as Fig. 1.

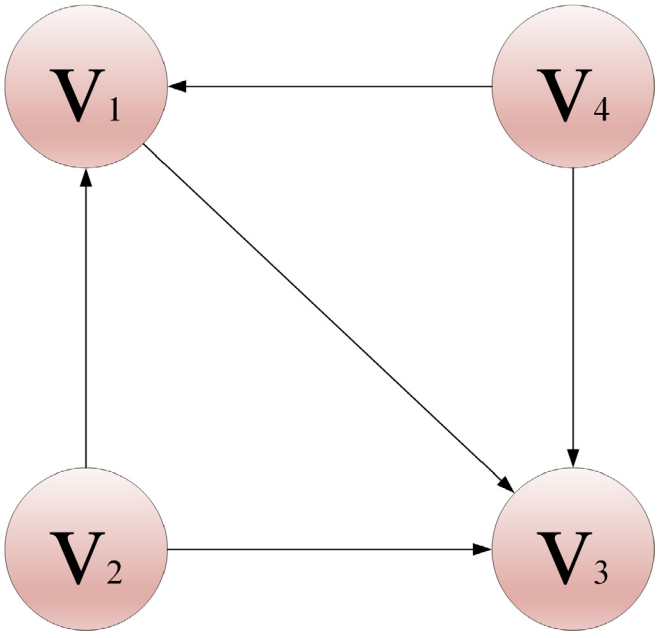

Fig. 1. The directed graph G.

The directed graph: The graph described as a mathematical formula:
$G = (V, E), V = n, E = m,$ and $V = \{v_1, v_2, \cdots, v_n\}, E = \{ < v_i, v_j > | v_i, v_j \in V \ , \ i \neq j$ Where, n is the number of vertices, m is the number of edges, and the order pair $< v_i, v_j >$ represents a directed edge of the vertex v_i to v_j. The directed graph model describes the interrelationship between the elements of the system.

Adjacency matrix: The relationship between two nodes in a directed graph G can be represented by an n × n adjacency matrix $A = (a_{ij})$.

$$a_{ij} = \begin{cases} 1, & < v_i \ , \ v_j > \in E \ , \ 1 \leq i \ , j \leq n \\ 0, & other \end{cases} \tag{1}$$

Where, n is the number of constituent elements. When the element v_i has no effect or no relation to v_j, a_{ij} is 0, if the element v_i has an effect or relationship with v_j, a_{ij} is 1.

Reachable matrix: The vertex v_i to the vertex v_j is reachable when there is a directional path $v_i = v'_0 | v'_1, \cdots, v'_m = v_j$. And an n × n reachable matrix $R = (r_{ij})$ is defined in the directed graph G.

$$r_{ij} = \begin{cases} 1, When \ v_i \ \ to \ v_j \ \ reach \ , \ 1 \leq i \ , j \leq n \\ 0, \ \ other \end{cases} \tag{2}$$

The nodes of the directed graph G are adjacent to the matrix A and the reachable matrix R.

$$A = \begin{vmatrix} v_1 & v_2 & v_3 & v_4 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{vmatrix} \quad R = \begin{vmatrix} v_1 & v_2 & v_3 & v_4 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{vmatrix}$$

The exponentiation is performed on the matrix $A + I$, and the exponentiation is obtained based on the Boolean algebraic operation (i.e. $0 + 0 = 0, 0 + 1 = 1, 1 + 1 = 1, 1 \times 1 \ 1, 1 \times 0 = 0 \times 1 = 0$) until the following Eq. (3) is accepted.

$$R \equiv (A + I)^{n+1} = (A + I)^n \neq \cdots \neq (A + I)^2 \neq A + I \tag{3}$$

The matrix $R = (A + I)^{n+1}$ is the reachable matrix.

The basic working principle of the ISM is shown in Fig. 2.

The main working procedure of the ISM is divided into the following steps:

Step 1: Set up the implementation of the ISM working group.

Step 2: Resolved the group set and discuss the content.

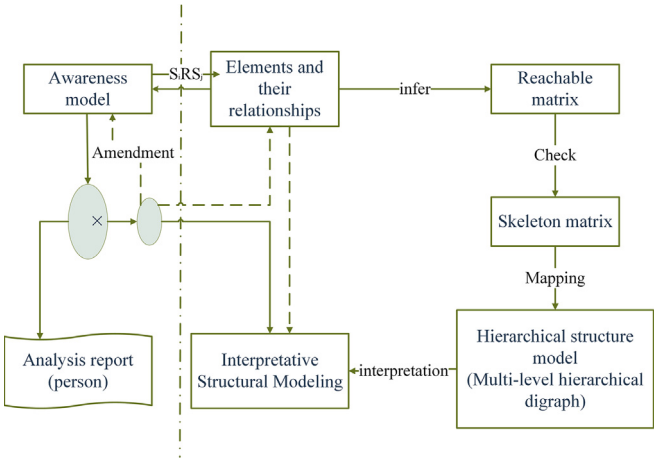

Fig. 2. ISM working principle diagram.

Step 3: Build the adjacent matrix and calculate its reachable matrix according to the elements of the content.

Step 4: Decompose the reachable matrix to construct the structure analysis model.

Step 5: Analyze the structural analysis model according to the actual work.

3.1.2. The AHP based on the entropy weight

The AHP combines qualitative analysis with quantitative analysis. The determination of weights is obtained based on the relative importance of the various factors. The AHP model based on the entropy method has a strong ability of distinguishing indexes when the indexes are the same, which makes the weight distribution more objective and accurate [9] [41].

The hierarchical distribution relation matrix of all factors can be obtained based on the calculation of the ISM, and then the factors of each lever can be fused by the AHP with the following steps. Assume that $X = \{X1, X2, \cdots, Xn\}$ is the data which has been calculated by the ISM.

The correlation between relevant nodes is determined to construct the correlation matrix. The nodes of $k_{ij}(x)$ are denoted by $x_j(1), x_j(2), x_j(3), x_j(4)$.

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

When the second node coincides with the third node, the standard correlation function is expressed as Eq. (5):

$$k_{ij}(x) = \begin{cases} 0 & x \notin [x_j(1), x_j(4)] \\ \dfrac{x_{ij} - x_j(1)}{x_j(2) - x_j(1)} & x \in [x_j(1), x_j(2)] \\ \dfrac{x_j(4) - x_{ij}}{x_j(4) - x_j(3)} & x \in [x_j(2), x_j(4)] \end{cases} \tag{5}$$

Apply the correlation function to X, the information matrix is obtained as Eq. (6):

$$K_{n\times m} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1m} \\ k_{21} & k_{22} & \cdots & k_{21} \\ \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & \cdots & k_{nm} \end{bmatrix}$$

Matrix Normalization:

$$k'_{ij} = \left(k_{ij} - \overline{k_j}\right) \Big/ S_j (\mathrm{i} = 1, 2, 3, \cdots, \mathrm{n}; \mathrm{j} = 1, 2, 3, \cdots \mathrm{m}) \tag{6}$$

Where,

$$\overline{k_j} = \frac{1}{n} \sum_{i=1}^{n} k_{ij} (j = 1, 2, \cdots, m) \tag{7}$$

$$R^j_{n\times m} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{21} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}$$

The orthogonal matrix $\mathrm{COR} = \mathrm{R}R^T$ is obtained by using $R^j_{n\times m}$.

The difference coefficient of the i $(\mathrm{i} = 1, \cdots, \mathrm{m})$ evaluation index can be calculated by using the entropy method as shown in Eq. (8).

$$E_i = 1 + k \sum_{j=1}^{n} r_{ij} \ln r_{ij} \tag{8}$$

Where, $\mathrm{k} = 1/\ln m$.

The greater the difference degree between the indicators, the greater the amount of information and the discrimination on the indicators [42]. And the weight of the important indicators is calculated as Eq. (9):

$$W_i = \frac{E_i}{\sum_{i=1}^{n} E_i} \tag{9}$$

The weight W is used to obtain the fused input data $\widehat{X}$ as Eq. (10):

$$\widehat{X} = X^T \mathrm{W} \tag{10}$$

3.2. The ELM

Single hidden layer feedback neural network has two more prominent abilities. (1) It can directly fit out the complex mapping function from the training sample. (2) It can provide a model for a large number of natural or artificial phenomena that are difficult to handle with traditional classification parameters. However, the single hidden layer feedback neural network lacks a relatively fast learning method. The BP algorithm needs to update $\mathrm{n(L+1)+L(m+1)}$ values for each iteration, which takes much less time than the tolerated time. Therefore, The ELM is introduced, which is a single hidden layer neural network. And the ELM can randomly initialize the input weights and offsets to get the corresponding output weights. Once the input weights and hidden layer offsets are randomly determined, the hidden layer output matrix is uniquely determined. Therefore, training the hidden-layer neural network can be transformed into solving a linear system.

Suppose there is a sample $(X_i,\ t_i)$, where, $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T \in R^n$, $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T \in R^m$. The single hidden layer neural network can be expressed as Eq. (11):

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_i + b_i) = o_j, j = 1, \cdots, N \tag{11}$$

Where, g(x) is an activation function, the stimulus function taken in this paper is sigmoidal function (Sig), the sigmoidal function is defined by Eq. (12):

$$\mathrm{S(x)} = \frac{1}{1 + e^{-x}} \tag{12}$$

$W_i = [w_{i,1}, w_{i,2}, \cdots, w_{i,n}]^T$ is the input weight, β is the output weight, and b_i is the offset of the i-th hidden layer. $W_i \cdot X_i$ represents the inner product of W_i and X_i. The output error is expressed as Eq. (13):

$$\sum_{j=1}^{N} \| o_i - t_i \| = 0 \tag{13}$$

And then,

$$\sum_{i=1}^{L}\beta_i g(W_i\cdot X_i+b_i)=t_j, j=1,\cdots,N \tag{14}$$

$$\mathrm{H}\beta=\mathrm{T} \tag{15}$$

Where, H is the output of hidden layer nodes, β is the output of weight, and T is the output of desired.

$$\mathrm{H}(W_1,\cdots,W_L,b_1,\cdots,b_L,X_1,\cdots,X_L)=\begin{bmatrix} g(W_1\cdot X_1+b_1) & \cdots & g(W_L\cdot X_1+b_L)\\ \vdots & \ddots & \vdots\\ g(W_1\cdot X_N+b_1) & \cdots & g(W_1\cdot X_N+b_L)\end{bmatrix}_{N\times L} \tag{16}$$

$$\beta=\begin{bmatrix}\beta_1^T\\ \vdots\\ \beta_L^T\end{bmatrix}_{L\times m},\quad \mathrm{T}=\begin{bmatrix}T_1^T\\ \vdots\\ T_L^T\end{bmatrix}_{N\times m},$$

In order to train the single hidden layer neural network, $\widehat{W}_i$, $\widehat{b}_i$ and $\widehat{\beta}_i$ are obtained by Eq. (17).

$$\| H\left(\widehat{W}_i,\ \widehat{b}_i\right)\widehat{\beta}_i-T\|=\min_{W,b,\beta}\|H\left(\widehat{W}_i,\ \widehat{b}_i\right)\widehat{\beta}_i-T\|, (\mathrm{i}=1,\cdots,\mathrm{L}) \tag{17}$$

To minimize the loss function

$$\mathrm{E}=\sum_{j=1}^{N}\left(\sum_{i=1}^{L}\beta_i g(W_i\cdot X_j+b_i)-t_j\right)^2 \tag{18}$$

Once the input weight W_i and the bias b_i are randomly determined, the output matrix H of the hidden layer is uniquely determined. A single hidden layer neural network can be transformed into solving a linear system $\mathrm{H}\beta=\mathrm{T}$. And then, the output weight β can be determined as Eq. (19)

$$\widehat{\beta}=H^{\dagger}T \tag{19}$$

where, $H^{\dagger}$ is the Moore − Penrose generalized inverse matrix of H. And the norm of solution β is the smallest and unique.

The process of the ELM is summarized as below:

Inputs: the training set: $\{x_j,t_j\}_{j=1}^{N}$, the Numbers of hidden neurons L and the activation function $g(\cdot)$.
Output: The output weights β.
Step1: Randomly generate $w_i,b_i,i=1,\cdots,L$;
Step 2: Calculate the response matrix of hidden layers H.
Step 3: Calculate the output weight β based on the response matrix H.

3.3. The framework of the ELM integrated ISM-AHP

Step 1: The training sample $X_i=[x_{i1},x_{i2},\cdots,x_{in}]^T\in R^n, i=1,2,\cdots,k$ are layered by using the ISM, which can intuitive clearly reflect the structure of the relationship between the input elements. Therefore, the multi-level hierarchical relation matrix of k factors is obtained. And the complex elements are divided into $\mathrm{m}(\mathrm{m}\le\mathrm{k})$ classes.

Step 2: Each layer of the input data is fusion by the AHP based on entropy weights respectively.

Step 3: The merged input data can be expressed as $X_j=[x_{j1},x_{j2},\cdots,x_{jn}]^T\in R^n, j=1,2,\cdots,m$. In this paper, a single output neural network structure is selected, and the target output vector is expressed as: $\mathrm{Y}=[y_{i1},y_{i2},\cdots,y_{in}]^T$.

Step 4: Initializing the ELM parameters. Meanwhile, the K-fold cross-validation method is applied to select the experimental data and test data set. And then the weights of the output are calculated.

Step 5: Calculating the generalization error and the training error of the ELM integrated ISM-AHP. The K-fold cross-validation method cuts the data into subsets, and subsets of the data are alternately trained to calculate the average of Mses.

Step 6: Return to step 5 to make the error measurement more accurate.

The flow chart of the ELM integrated ISM-AHP is shown in Fig. 3.

4. UCI test validation

UCI Machine Learning Repository is a commonly standard test data set, which can be used for classification, regression, clustering and recommending the system task. Meanwhile, the UCI database library currently has 378 data sets. And each dataset has specific attributes and categories and different sizes. In order to verify the effectiveness and accuracy of the proposed model, this paper extracted the several groups of UCI data sets [43,44] as shown in Table 1.

Firstly, the ISM is used to layer the input data. UCI dataset samples, the breast-cancer-Wisconsin-I, the letter-recognition and the iris are levered into 3, 2, and 3 layers, respectively (The layer numbers of the ISM is equal to the input of the ELM integrated ISM-AHP). Therefore, the AHP is applied to fuse each layer separately. And then, the experimental parameters initialization of these models with BP, RBF, ELM, BP integrated ISM-AHP, the RBF integrated ISM-AHP and the ELM integrated ISM-AHP respectively are

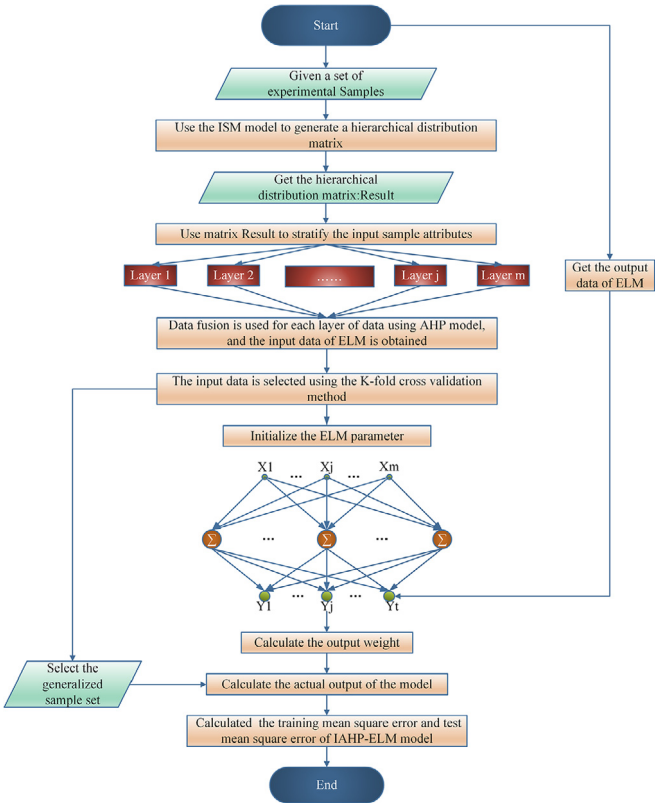

Fig. 3. The flow chart of the ELM integrated ISM-AHP.

Table 1
UCI data sets samples.

Data Sets	Samples		Attributes	
	Training	Testing	Inputs	Outputs
breast-cancer-Wisconsin-I	174	20	34	1
letter-recognition	18000	2000	16	1
Iris	135	15	4	1

shown in Table 2.

The sigmoidal function of the ELM is chosen as the excitation function. By comparing the experimental Mse to assess the quality of the experimental results, the Mse is calculated as Eq. (20).

$$\mathrm{Mse} = 1/\mathrm{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2 \tag{20}$$

Where, y_i represents the actual value, $\widehat{y_i}$ represents the predicted value, and the range is from 1 to n. The Mse can evaluate the degree of data change, and the Mse value is closer to 0, indicating that the predictive model describes the experimental data with better accuracy. The results are shown in Tables 3–5. The training mean square error and the test mean square error is expressed as Mse_train and Mse_test, respectively.

In the breast-cancer-Wisconsin-I experiment, the results of the ELM integrated ISM-AHP and the ELM models are clearly the best. Meanwhile, compared with the ELM, the Mse_train and Mse_test of the ELM integrated ISM-AHP is reduced by 0.0036 and 0.002, respectively. In the letter experiment, the Mse_train and the Mse_test is reduced by 0.0023 and 0.0015, respectively. In the iris experiment, the Mse_test is reduced by about 50% compared to the ELM and the relatively well-behaved the BP and the RBF. Moreover, the number of hidden layer nodes in ELM integrated ISM-AHP is 2. The results show that the efficiency and accuracy of the proposed method is verified.

5. Case study: production prediction and energy-saving of complex chemical processes

In the complex chemical processes, ethylene is an important indicator of national energy capacity and PTA is one of the important organic raw materials. Therefore, how to increase the production of ethylene and PTA to reduce energy consumption has been a major concern. The proposed model is used to predicate the production and save energy of the ethylene chemical process and the PTA chemical process.

5.1. Production predication and energy-saving for the ethylene industry

A typical flowchart of the ethylene plant is as shown in Fig. 4. The main process of ethylene production consists two parts: cracking and separation. The separation section is a three-step process which can be described as rapid cooling, compression and final separation [45]. The production and operation of the cracking furnace consumes a large amount of fuel, which results in huge energy usage in ethylene production together with other high-energy-consumption taches.

The experiment data sources come from the industrial data of two ethylene plants with two different kinds of technology under 800,000 Tons scale in the ethylene production system from 2008 to 2013 [7]. The ratio of raw materials and composition is very

Table 2
Experimental parameters initialization.

	ELM integrated ISM-AHP	ELM	RBF integrated ISM-AHP	RBF	BP integrated ISM-AHP	BP
Maximum number of iterations	–	–	–	–	–	5000
Default accuracy	–	–	1e-3	1e-3	1e-3	1e-3
Momentum factor	–	–	–	–	–	0.9
Incentive function	Sig	Sig	–	–	–	–

Table 3
Test results of the breast-cancer-Wisconsin-I standard data set.

	ELM integrated ISM-AHP	ELM	RBF integrated ISM-AHP	RBF	BP integrated ISM-AHP	BP
Hidden nodes number	2	8	4	4	8	8
Mse_train	0.0115	0.0481	0.6866	0.3911	0.7622	0.6989
Mse_test	0.0033	0.0053	0.7119	0.7429	0.6776	0.7283

Table 4
Test results of the letter standard data set.

	ELM integrated ISM-AHP	ELM	RBF integrated ISM-AHP	RBF	BP integrated ISM-AHP	BP
Hidden nodes number	2	8	4	4	8	8
Mse_train	0.0023	0.2735	0.3518	0.2501	0.3733	0.2639
Mse_test	0.0015	0.2724	0.3519	0.2505	0.3739	0.2636

Table 5
Test results of the iris standard data set.

	ELM integrated ISM-AHP	ELM	RBF integrated ISM-AHP	RBF	BP integrated ISM-AHP	BP
Hidden nodes number	2	8	4	4	8	8
Mse_train	0.0133	0.2681	0.1364	0.042	0.1512	0.0375
Mse_test	0.0407	0.0667	0.1429	0.0471	0.1587	0.042

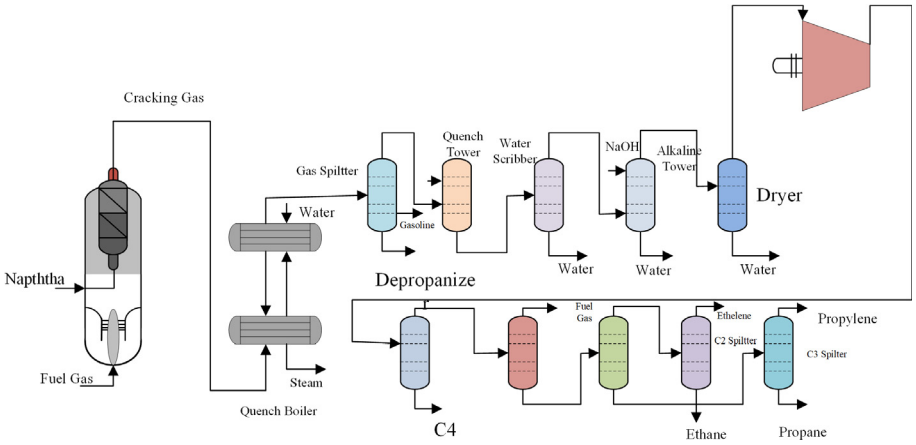

Fig. 4. A typical flowchart of the ethylene plant.

complicated, the factors which related to energy efficiency mainly include: (I) the raw material; (II) the consumption of fuel, steam, water and electricity; (III) the product of the plants. As the raw material, crude oil includes C345, hydrogenation tail oil, light diesel fuel, naphtha and others. The fuel includes light weight oil, heavy weight oil and fuel gas. The consumption of steam is divided into super-high-pressure steam, high-pressure steam, medium-pressure steam and low-pressure steam. The water consists of circulating water, industrial water, boiler water and other water. The relationship between ethylene production and the factors is as shown in Fig. 5. The final ethylene yield (unit: T) is indicated by a broken line. The units of the raw material are Ton, and the units of other inputs are all converted to GJ/Ton ethylene [8,9].

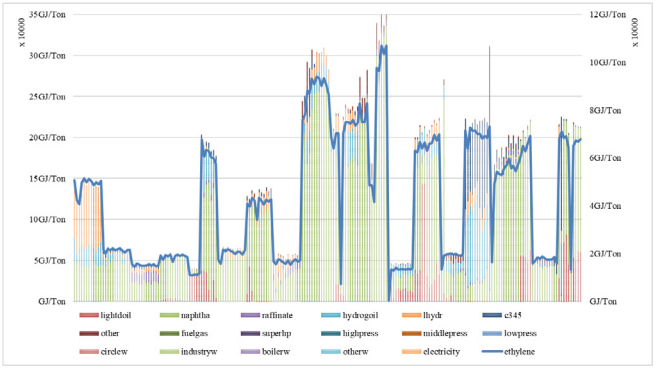

Fig. 5. The relationship between ethylene production and raw material ratio.

The partial correlation coefficient threshold of the ISM is set to 0.5 [46], and the input of the seventeen categories of ethylene is divided into three layers by using the ISM as shown in Fig. 6. As is shown in the first layer has seven attributes, the second layer has seven attributes and the bottom layer has three attributes. And then, the weights of each layer are obtained by the AHP as shown in Fig. 7. Each layer can be fused into one input based on the

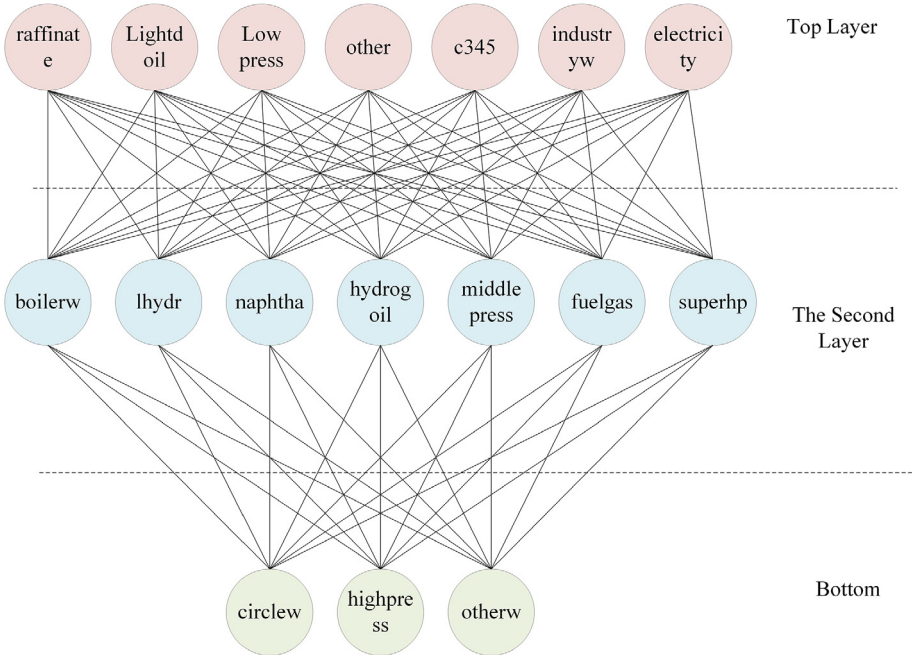

Fig. 6. The hierarchical relationship among the input factors of the ethylene industry.

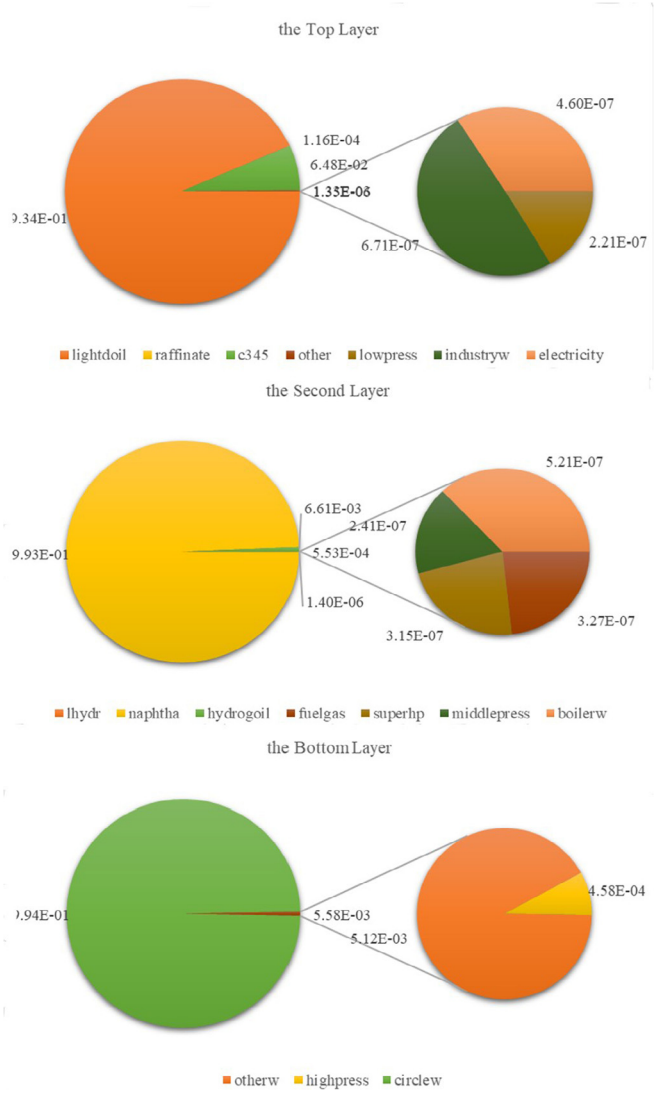

Fig. 7. The distribution of weights obtained by AHP fusion of input properties in the ethylene industry.

corresponding weight matrix. Therefore, the 17 inputs are reduced to 3 inputs. We selected 190 sets of actual statistical data for modeling and predicted 22 sets of ethylene plant output. At the same time the above experiment data is modeled by the BP, the RBF, the ELM, the BP integrated ISM-AHP and the RBF integrated ISM-AHP to compare the forecasting performance with the proposed method. Therefore, the initial values of six kinds of network parameters are shown respectively in Table 6.

The Mse_train and the Mse_test is obtained by these neural networks is shown in Fig. 8, respectively. Compared with the other

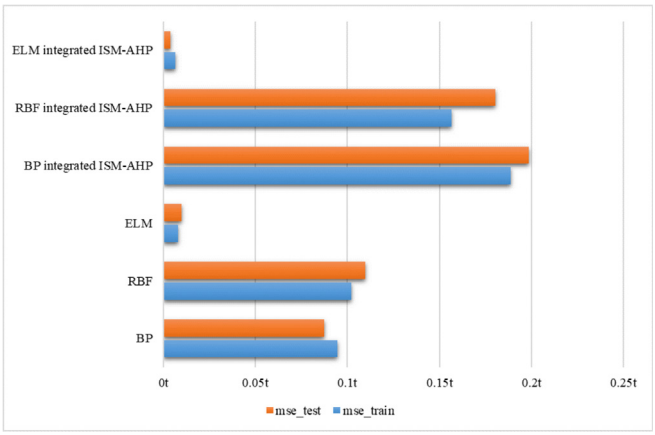

Fig. 8. The Mses for different networks in ethylene prediction.

neural network models, the Mses of the proposed method are the lowest. The Mse_train is changed from 0.0231 to 0.0209 while the Mse_test is reduced from 0.0077 to 0.0038. Meanwhile, the data shows that the generalization ability of the proposed method is obviously better than the others.

The predicted yields obtained by the proposed method and the actual yields of ethylene production processes are shown in Fig. 9.

The adjustment of production parameters of 3th and 4th sample point is as shown in Fig. 10. The actual ethylene production output of the 3th sample point is larger, which means that the plant operates without full load in the ethylene manufacturing process and the energy efficiency is low. According to the weights obtained from the AHP. The factors that have a great influence on the result will be adjusted to realize the effective production of the plant. If the ethylene production inputs of the 3th reduce 0.02 GJ fuel gas,0.07 GJ electricity, 13016t naphtha and increase 12202t light oil, 50 GJ recycled water, 0.18 GJ industrial water/per ton of ethylene, and the outputs increase 2739.54 Ton ethylene, then the efficiency value can achieve the effective level. Conversely, the actual ethylene production output of the 4th sample point is lower than the predicted value, which means that the plant production is overloaded. If its ethylene production inputs reduce 8043 light oil, 0.07 GJ fuel gas, 25 GJ recycled water, 0.03 GJ electricity and increase 0.12 GJ industrial water/per ton of ethylene, then the efficiency value can achieve the effective level.

The consumption of acetic acid is an important indicator to measure the advancement and effectiveness of PTA technology. And the most important way to reduce the consumption of acetic acid is the optimal control of the solvent system. Solvent system can be divided into three parts: the solvent dehydration tower, the re-boiler and the reflux tank. The schematic flow diagram of a solvent dehydration tower is shown in Fig. 11 [7,45].

The experiment data sources come from the PTA solvent system during two months of 2011 [9]. The main factors affecting the PTA

Table 6
The parameter initialization of different Models.

	ELM integrated ISM-AHP	ELM	RBF integrated ISM-AHP	RBF	BP integrated ISM-AHP	BP
The number of nodes in the input layer	3	17	3	17	3	17
The number of nodes in the output layer	1	1	1	1	1	1
Maximum number of iterations	–	–	–	–	–	5000
Default accuracy	–	–	1e-3	1e-3	1e-3	1e-3
Momentum factor	–	–	–	–	–	0.9
Incentive function	Sig	Sig	–	–	–	–

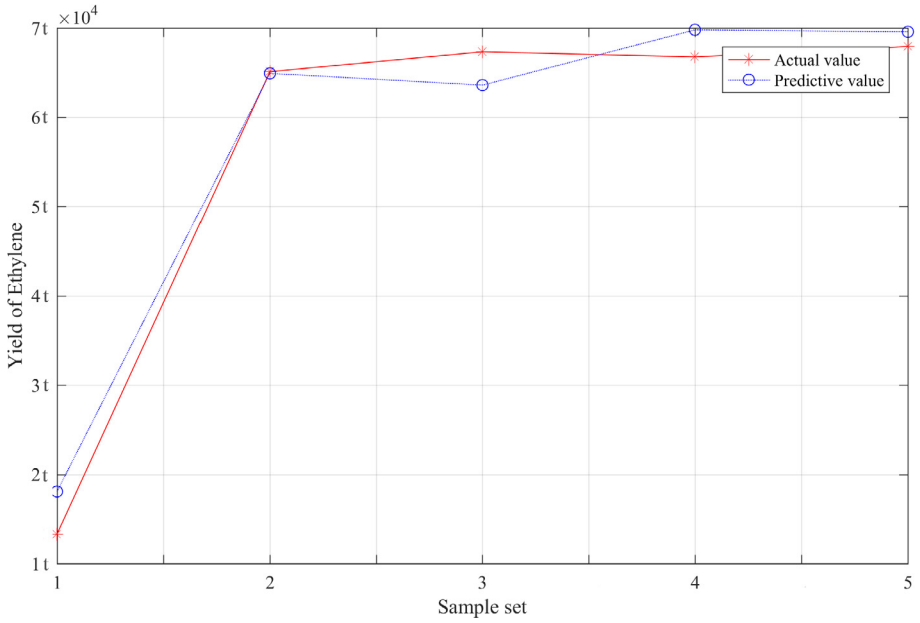

Fig. 9. Comparison of predicted and actual values in ethylene production.

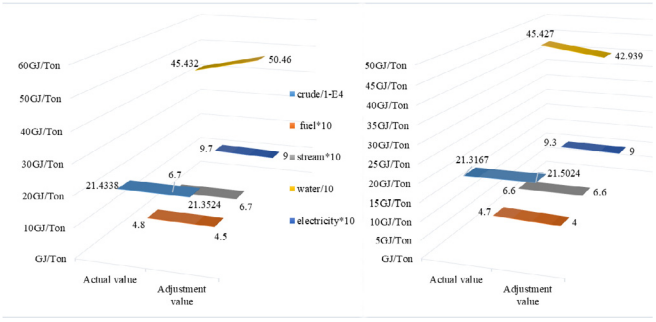

Fig. 10. The adjustment of production parameters of 3th and 4th sample point. Production predication and energy-saving of PTA industry.

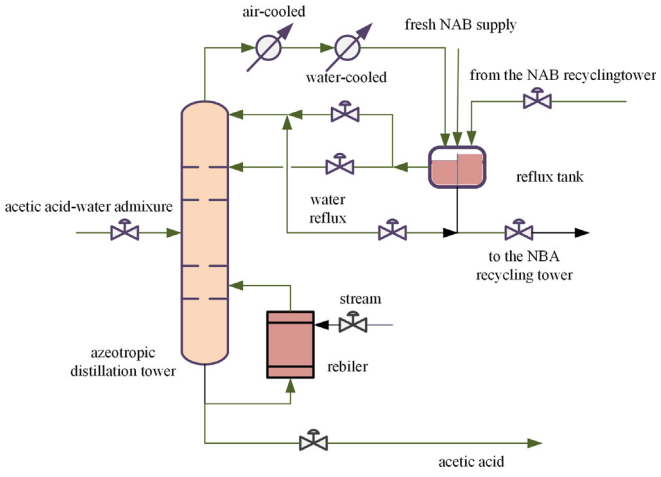

Fig. 11. The schematic flow diagram of a solvent dehydration tower.

solvent system are shown in Table 7. The following 17 variables are used as the input, and the PTA yield is the only output. In order to optimize the input parameters, the experimental sample set

Table 7
The input variable and noun abbreviation of the PTA plant.

No.	Input Variable	ELM
1	feed composition	FC
2	feed quantity	FC1501
3	water reflux	FC1502
4	NBA main reflux	FC1503
5	NBA side reflux	FC1504
6	steam flow	FC1507
7	produced quantity of the top tower	FI1511
8	feed temperature	TI1504
9	reflux temperature	TI1510
10	temperature of the top tower	TI1511
11	temperature point above the 35th tray	TI1515
12	temperature point between the 35th tray and the 40th tray	TI1516
13	temperature point between the 44th tray and the 50th tray	TI1517
14	tray temperature near the up sensitive plate	TC1503a
15	tray temperature near the low sensitive plate	TC1503b
16	temperature point between the 53rd tray and the 58th tray	TC1501
17	reflux tank level	LC1503a

includes a set of 260 industrial data. After the data selection by the K-fold cross-validation, 234 data are used for training and 26 data are used for generalization. Finally, the averages of the Mses are

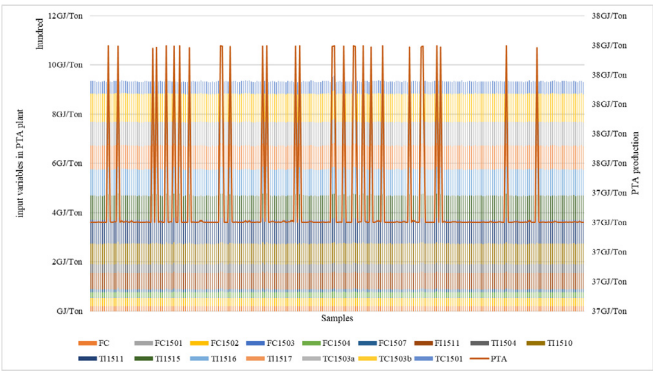

Fig. 12. Relationship between PTA yield and input variables.

Table 8
Parameter initialization of the proposed method.

Parameter	Value
The number of hidden nodes	4
The number of nodes in the input layer	3
The number of nodes in the output layer	1
Incentive function	Sig
partial correlation coefficient threshold of ISM	0.5

taken as the final result of the experiment. The relationship between PTA production and the factors is shown in Fig. 12.

In the PTA chemical process, the parameters of the proposed method are initialized as Table 8. The 17 input properties of PTA are analyzed by the ISM and the hierarchical structure is shown in Fig. 13.

In the PTA chemical process, the PTA training inputs are reduced from 17 elements to 3 elements. As is shown in the first layer has nine attributes, the second layer has six attributes and the bottom has two attributes. According to the sample set provided by the experiment, the distribution weights are obtained as shown in Fig. 14.

It is known from the above analysis that the model based on the proposed method can reach higher generalization accuracy during treating with the complicated PTA solvent system data, and increase the accuracy for predicting the PTA industrial capacity. The comparison of predicted results among the BP, the RBF, the ELM, the BP integrated ISM-AHP, the RBF integrated ISM-AHP and the ELM integrated ISM-AHP network is shown in Fig. 15. Both of the ELM and the ELM integrated ISM-AHP perform well. However, the Mse_train and the Mse_test of the ELM integrated ISM-AHP decreased by 0.0022 and 0.0039, respectively.

The predicted yields obtained by the proposed method and the actual yields of PTA production processes are shown in Fig. 16. In order to realize the energy savings, the adjustment of production parameters (TI1511, TC1501, TI1517, TI1516, TC1503b, TC1503a) of the 4th and the 5th sample points are described in Fig. 17.

It can be seen from Figs. 16 and 17 that the actual PTA production output of the 5th sample point is larger, which means that the plant operates without full load. However, the actual PTA production output of the 4th sample point is lower than the predicted value, which means that the plant production is overloaded. Therefore, the parameters TI1511, TI1516 and TI1517 of the 4th should be

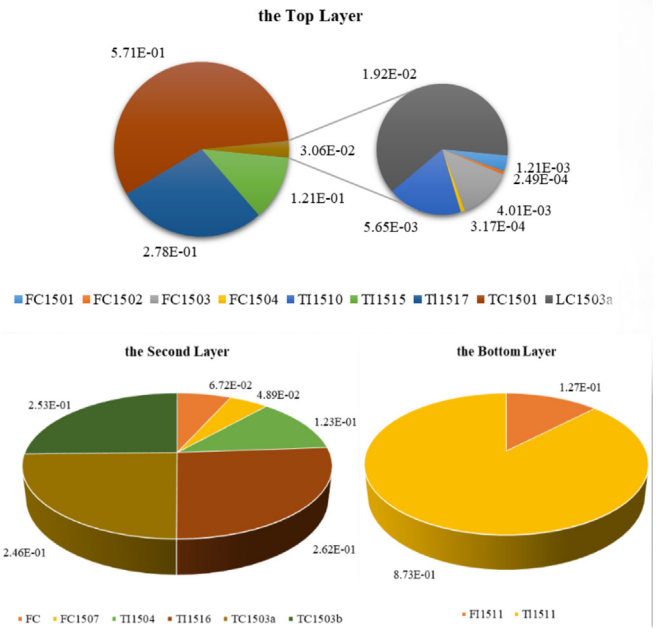

Fig. 14. The distribution weights obtained by the AHP in the PTA industry.

reduced 0.01291, 1.53165, and 0.66638, respectively. And the parameters of TC1503a and TC1503b should be increased 1.309866 and 0.674493, and the consumption of acetic acid reduces 0.012. Meanwhile, the parameter TI1511 of the 5th should be reduced 0.08292. And TC1501, TI1517, TI1516, TC1503a and TC1503b should be increased 0.008927, 0.438599, 4.23252, 0.309899 and 0.114916, respectively. Meanwhile, the consumption of acetic acid increase 0.005. Based on the adjustment of production parameters, the efficiency value of the 4th and the 5th sample point can achieve the effective level.

6. Discussion

First, the ELM based on the ISM-AHP is proposed. The factors that affect the productivity are divided into different levers by the ISM. And then the attributes of each layer are fused by the AHP

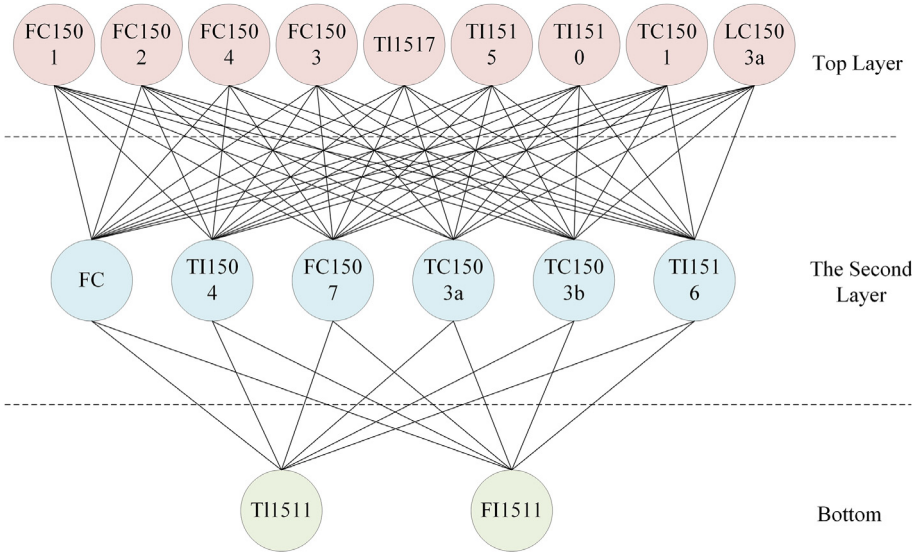

Fig. 13. The hierarchical relationship among the input factors of the PTA industry.

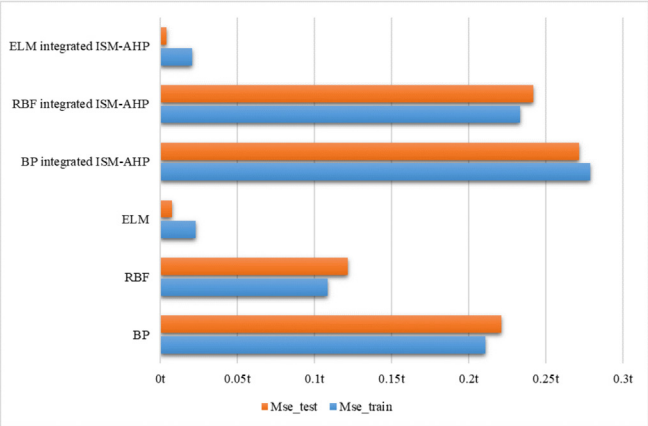

Fig. 15. The Mses for different networks in PTA prediction.

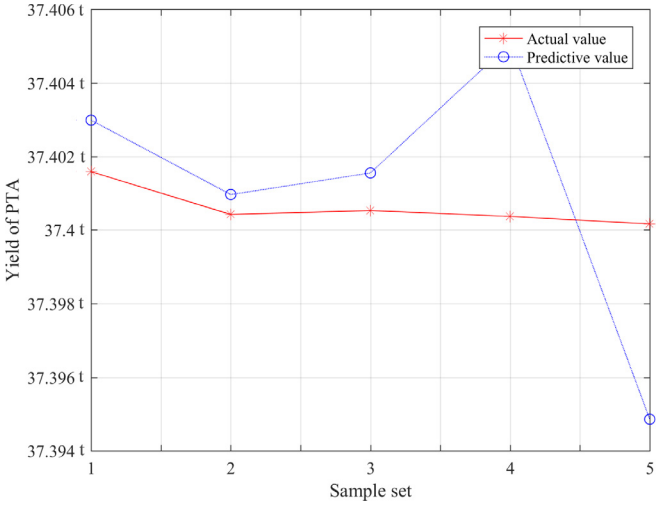

Fig. 16. Comparison of predicted and actual values in PTA production.

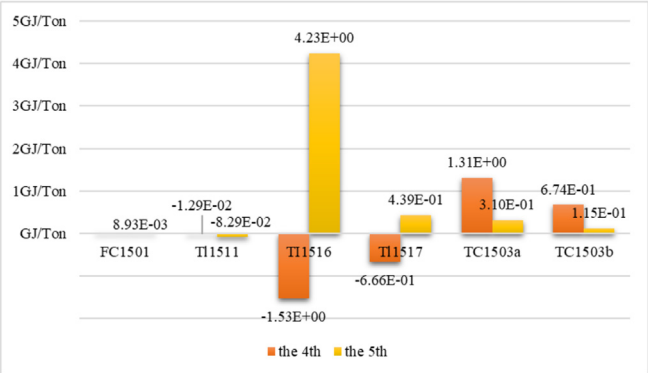

Fig. 17. The adjustment of production parameters of the 4th and the 5th sample point.

based on the entropy weight, which greatly reduces the complexity of the input attributes as the inputs of the ELM. Therefore, compared with the ELM, the RBF and the BP, the validity and the practicability of the proposed method are validated by three UCI datasets.

Second, the proposed method is applied in the production prediction and energy-saving of the ethylene production system and the PTA production system in complex chemical processes. Compared with the BP, the RBF, the ELM, the BP integrated ISM-AHP and the RBF integrated ISM-AHP, the prediction accuracy of the ethylene and the PTA production reaches about 99%. Meanwhile, we can maximize the production with the least energy consumption to improve the energy efficiency and save the energy.

Third, the proposed model still has some drawbacks. Compared with the ELM model, the accuracy of the proposed method is not very obvious. Moreover, more in-depth researches should be done on adjusting the threshold of the ISM. Therefore, we can take a sliding window method to automatically adjust the threshold of the ISM. Meanwhile, we will use the proposed model to dispose the data of per second or per minute to achieve the real time application and compare with the ELM model.

7. Conclusion

This paper proposes a production prediction model based on the ELM integrated the ISM and the AHP, which is used in complex chemical industries. The main attributes are obtained and merged by using the ISM and the AHP based on the entropy weight to build the production capacity prediction and energy-saving model based on the ELM. Compared with the ELM, the RBF and the BP, the validity and practicability of the proposed method are verified by UCI datasets. Finally, the proposed method is used in the production prediction and energy-saving of ethylene production system and PTA production system in complex chemical processes. Compared with the BP, the RBF, the ELM, the BP integrated ISM-AHP and the RBF integrated ISM-AHP, the experimental results showed the proposed method can reduce the number of hidden layer nodes, and improve the training time and effectively process the high-dimensional data. Meanwhile, the prediction accuracy of the ethylene and the PTA production reaches about 99% to guide the raw materials allocation and improve the energy efficiency.

In our further works, the pollutant emissions should be taken into account. Meanwhile, we will integrate some artificial intelligence methods, such as particle swarm optimization, deep learning etc., to adaptively adjust the threshold of the ISM. Furthermore, the proposed model can be applied widely in energy-saving and carbon emission reduction of other complex industries.

Acknowledgments

This research was partly funded by National Key Research and Development Program of China (2018YFB0803501), National Natural Science Foundation of China (61603025, 61533003 and 61374166) and the Natural Science Foundation of Beijing, China (4162045).

References

[1] Geng ZQ, Wang ZK, Zhu QX, Han YM. Energy efficiency evaluation method based on IDA-DEA and its applications in ethylene industries. CIESC J 2017;66(3):910–5.
[2] Yu RW, Ma GF, Xu YH. Ethylene business review of China petrochemical in 2015. Ethyl Ind 2016;28(1):1–6.
[3] report2016–2021 China ethylene production equipment industry market demand and investment advisory report. China Institute of Industry.
[4] Neelis M, Patel M, Bach P, et al. Analysis of energy use and carbon losses in the chemical industry. Appl Energy 2007;84(7):853–62.
[5] Lei L, Qian ZM. Domestic capacity expanding design and innovation of PTA plant oxidation reactor. Modern Chemical Industry; 2014.
[6] Garg KK, Prasad B, Srivastava VC. Comparative study of industrial and laboratory prepared purified terephthalic acid (PTA) waste water with electrocoagulation process. Separ Purif Technol 2014;128(19):80–8.
[7] Geng ZQ, Qin L, Han YM, Zhu QX. Energy-saving and prediction modeling of petrochemical industries: a novel ELM based on FAHP. Energy 2017;122:350–62.
[8] Wang J, Feng L. Curve-fitting models for fossil fuel production forecasting: key influence factors. J Nat Gas Sci Eng 2016;32:138–49.

[9] Geng ZQ, Yang X, Han YM, Zhu QX. Energy optimization and analysis modeling based on extreme learning machine integrated index decomposition analysis: application to complex chemical processes. Energy 2017;120:67–78.
[10] Han YM, Zhu QX, Geng ZQ, Xu Y. Energy and carbon emissions analysis and prediction of complex petrochemical systems based on an improved extreme learning machine integrated interpretative structural model. Appl Therm Eng 2017;115:280–91.
[11] Betiku E, Odude VO, Ishola NB, et al. Predictive capability evaluation of RSM, ANFIS and ANN: a case of reduction of high free fatty acid of palm kernel oil via esterification process. Energy Convers Manag 2016;124:219–30.
[12] Chen Y, Hao YA. Feature weighted support vector machine and K-Nearest neighbor algorithm for stock market indices prediction. Expert Syst Appl 2017;80:340–55.
[13] Logo GA, Zilio C, Ortombina L, et al. Application of Artificial Neural Network (ANN) for modeling oxide-based nanofluids dynamic viscosity. Int Commun Heat Mass Tran 2017;83:8–14.
[14] Pereira Claudio MNA, Schirru Roberto, Gomes Kelcio J, Cunha Jose L. Development of a mobile dose prediction system based on artificial neural networks for NPP emergencies with radioactive material releases. Ann Nucl Energy 2017;219(7).
[15] Geng ZQ, Gao HC, Wang YQ, Han YM, Zhu QX. Energy saving analysis and management modeling based on index decomposition analysis integrated energy saving potential method: application to complex chemical processes. Energy Convers Manag 2017;145:41–52.
[16] Rumelhart DE, Hinton GE, Williams RJ. Learning representation by back-propagating errors. Nature 1986;323(3):533–6.
[17] Dang X, Yan L, Jiang H, et al. Open-circuit voltage-based state of charge estimation of lithium-ion power battery by combining controlled auto-regressive and moving average modeling with feedforward-feedback compensation method. Int J Electr Power Energy Syst 2017;90:27–36.
[18] Hardy RL. Multiquadric equations of topography and other irregular surfaces. J Geophys Res 1971;76(8):1905–15.
[19] Rezaei J, Shahbakhti M, Bahri B, et al. Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks. Appl Energy 2015;138(C):460–73.
[20] Dong L, Liu XJ, Yang YH. Investigation of uncertainty quantification method for BE models using MCMC approach and application to assessment with FEBA data. Ann Nucl Energy 2017;107:62–70.
[21] Fang Y, Fei J, Ma K. Model reference adaptive sliding mode control using RBF neural network for active power filter. Int J Electr Power Energy Syst 2015;73:249–58.
[22] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE international joint conference on neural networks, 2004. Proceedings, vol. 2. IEEE; 2005. p. 985–90.
[23] Naji S, Keivani A, Shamshirband S, et al. Estimating building energy consumption using extreme learning machine method. Energy 2016;97:506–16.
[24] Li X, Xie H, Wang R, et al. Empirical analysis: stock market prediction via extreme learning machine. Neural Comput Appl 2016;27(1):67–78.
[25] Wong PK, Wong KI, Chi MV, et al. Modeling and optimization of biodiesel engine performance using kernel-based extreme learning machine and cuckoo search. Renew Energy 2015;74(C):640–7.
[26] Warfield JN. Developing interconnection matrices in structural modeling. In: Systems man & cybernetics IEEE transactions on. SMC; 2010. p. 81–7. 4(1).
[27] Gao H, Xu Y, Zhu Q. Spatial interpretive structural model identification and AHP-based multimodule fusion for alarm root-cause diagnosis in chemical processes. Ind Eng Chem Res 2016;55(12).
[28] Dong D, Liu J, Zhou H. Influence factor analysis of supply chain resilience using ISM. In: International conference on service systems and service management. IEEE; 2016. p. 1–5.
[29] Liu Z, Shen X. Analysis of the influence factors of college students employment based on the interpretative structural model. In: International conference on advances in energy, environment and chemical engineering; 2016.
[30] Zhang C, Sun L, Wen F, et al. An interpretative structural modeling based network reconfiguration strategy for power systems. Int J Electr Power Energy Syst 2015;65(65):83–93.
[31] Re J, Tan S, Goodsite ME, et al. Sustainability, shale gas, and energy transition in China: assessing barriers and prioritizing strategic measures. Energy 2015;84:551–62.
[32] Zeng WM, Liu DN, Hu Y, et al. Analysis of low carbon influence factors in power industry based on the ISM. Appl Mech Mater 2014;675–677:1721–6.
[33] Saaty TL. A scaling method for priorities in hierarchical structures. J Math Psychol 1977;15(3):234–81.
[34] Haddad B, Liazid A, Ferreira P. A multi-criteria approach to rank renewables for the Algerian electricity system. Renew Energy 2017;107:462–72.
[35] Sidhu S, Nehra V, Luthra S. Investigation of feasibility study of solar farms deployment using hybrid AHP-TOPSIS analysis: case study of India. Renew Sustain Energy Rev 2017;73:496–511.
[36] Wang L, Sharkh S, Chipperfield A, et al. Dispatch of vehicle-to-grid battery storage using an analytic hierarchy process. IEEE Trans Veh Technol 2017;66(4):2952–65.
[37] Chen T, Shen D, Jin Y, et al. Comprehensive evaluation of environ-economic benefits of anaerobic digestion technology in an integrated food waste-based methane plant using a fuzzy mathematical model. Appl Energy 2017;208:666–7.
[38] Santos LFDOM, Osiro L, Lima RHP. A model based on 2-tuple fuzzy linguistic representation and Analytic Hierarchy Process for supplier segmentation using qualitative and quantitative criteria. Expert Syst Appl 2017;79:53–64.
[39] Taylan O, Kaya D, Demirbas A. An integrated multi attribute decision model for energy efficiency processes in petrochemical industry applying fuzzy set theory. Energy Convers Manag 2016;117:501–12.
[40] Qiu ZY, Zhang LL, Yang DS, et al. Application of interpretative structural modeling method on naval vessels training Program. Fire Control Command Control 2013;38(4):21–5.
[41] Yan D, Zhang SB, Zong K, et al. Identification of key nodes in a complex network based on AHP-entropy method. J Guangxi Univ (Philos Soc Sci) 2016;41(6):1933–9.
[42] Yan D, Zhang SB, Zong K, et al. Key node recognition method for complex networks based on AHP-entropy weight method. Journal of Guangxi University (Natural Science Edition) 2016;41(6):1933–9.
[43] Kim T, Chung BD, Lee JS. Incorporating receiver operating characteristics into naive Bayes for unbalanced data classification. Computing 2016;99(3):1–16.
[44] Ayed AB, Halima MB, Alimi AM. Cluster forests based fuzzy C-Means for data clustering. In: International conference on computational intelligence in security for information systems; 2016.
[45] Geng ZQ, Yang X, Han YM, Zhu QX. Energy optimization and analysis modeling based on extreme learning machine integrated index decomposition analysis: application to complex chemical processes. Energy 2017;120:67–78.
[46] Han YM, Geng ZQ, Gu XB, Zhu QX. Energy efficiency analysis based on DEA integrated ISM: a case study for Chinese ethylene industries. Eng Appl Artif Intell 2015;45:80–9.