Energy 142 (2018) 400–410

Contents lists available at ScienceDirect

Energy

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

# A novel self-organizing cosine similarity learning network: An application to production prediction of petrochemical systems

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

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

ARTICLE INFO

*Article history:*
Received 23 January 2017
Received in revised form
23 August 2017
Accepted 4 October 2017
Available online 6 October 2017

*Keywords:*
Neural network
Self-organizing
Cosine similarity
Entropy
Production prediction
Petrochemical systems

ABSTRACT

Single layer feed-forward network (SLFN) is well applied to find mapping relationships between the input data and the output data. However, the SLFN has two obvious shortcomings of the indetermination structure and parameters. Therefore, this paper proposes a novel self-organizing cosine similarity learning network (SO-CSLN), which can obtain a stable structure and suitable parameters. The hidden layer nodes of the SO-CSLN are determined by the rank of the sample covariance matrix based on the central limit theorem. And then the weights are obtained by the entropy theory and the cosine similarity theory. Moreover, compared with the SLFN, the proposed algorithm can overcome the shortcomings of the SLFN and provide better performance with faster convergence and smaller generalization error through different UCI data sets. Finally, the proposed method is applied to building the production prediction model of the ethylene production system in petrochemical industries. The experiment results show that the effectiveness and the practicality of the proposed method. Meanwhile, it can guide ethylene production and improve the energy efficiency.



## 1. Introduction

Energy is a material base of modern economy and the cornerstone of modern civilization. With the rapid development of economy, the energy efficiency improvement of industrial process cannot be separated from the high-quality energy issues and the advanced energy technologies. Ou et al. propose a method to apply a hybrid bird-mating optimization approach and a direct building algorithm for microgrid distribution (MGD) power flow analysis, serving as beneficial references for the improvement of electric power grid operations [1,2]. A novel unsymmetrical faults analysis method for microgrid (MG) is proposed [3]. Dynamic operation and control strategies for a microgrid hybrid wind–PV (photovoltaic) –FC (fuel cell) based power supply system proposed by Ou and Hong, analyzing the performance of the PV generation system [4]. Han et al. proposes a DEA-BP approach which could help to build a multi-input-multi-output energy optimization and prediction model [5]. Tc Ou et al. proposed a functional link-based novel recurrent fuzzy neural network (FLNRFNN). The result shows that the proposed controller can achieve better damping characteristics under unstable conditions [6]. Moreover, petrochemical systems make a difference in developing the industry development. And the production prediction is the main means to improving the energy efficiency of petrochemical industries. Nowadays, Artificial neural network (ANN) is widely used to solve the production prediction and the energy-saving in many industrial processes.

The ANN is a powerful tool that has been widely applied to model the nonlinear statistical data, make decision and build analysis and forecast model [7]. On the issue of the disease detection, the ANN is an ingenious triage tool, well applied to predict medical diagnosis and also be used to learn the unknown nonlinear conditions [8]. Chen et al. developed a neural network disturbance observer and put forward an integrated approach to calculate the workload based on the ANN [9]. Beşikçi et al. developed a decision support system (DSS) employing ANN-based fuel prediction model for on-board ship [10].

In 1986, Rumelhart put forward the back propagation neural network (BP network) [11]. The BP network has the strong nonlinear mapping ability to identify and classification capabilities

* Corresponding author. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China.
** Corresponding author. College of Information Science and 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).

https://doi.org/10.1016/j.energy.2017.10.017
0360-5442/

to input samples, a powerful nonlinear processing capability to perform better in the nonlinear classification problem and a capability of optimization. However, EZ Song et al. established the financial distinction model based on analyzing the BP network method and obtained the more precise forecast result [12]. In order to predict the position of the picking staff, Kou et al. developed a feature selection based on the BP network, which uses AIS to determine the connecting weights for the BP network [13]. Also the BP network is applied into predicting the production of field grown potatoes to solve a lot of complex situations [14]. Deosarkar et al. used the BP network model successfully for predicting the effective viscosity of magnetite ore slurries [15]. Nevertheless, the BP network still has the disadvantages of setting a lot of network training parameters and producing the local optima easily. In 1985, Powell proposed the multi-variable interpolation of radial basis function (RBF) method [16]. Hond et al. proposed a method used the RBF and the improved Elman neural network (ENN)for maximum power point tracking (MPPT) [17]. The results showed the RBF can effectively improve the robotic manipulators' performance. However, the RBF cannot explain the process and basis of reasoning, because it is easy to lose the data. Thus the learning algorithm need to be further improved and enhanced. In 2006, Huang proposed the extreme learning machine (ELM) [18]. The ELM does not need to adapt the input network weights or the bias hidden elements during the execution and produces the unique optimal solution with faster learning speed and better generalized performance [19]. The ELM has been widely used in energy analysis and the neural computation. Aghbashlo proposed the method based on the ELM to build the performance model of the diesel engine [20] and Li et al. raised an approach based on the ELM and improved the forecasting accuracy significantly [21]. Kariminia et al. presented the systematic ELM to analyze the thermal comfort at a public urban space [22]. Wang et al. proposed a novel NMF-based image quality assessment metric using the ELM, and the result showed that the ELM provided better generalization performance with much faster learning speed and less human intervene [23]. However, the ELM cannot determine the number of hidden layer nodes adaptively. The hidden layer nodes of the ELM needed to be determined by trying constantly. So that the optimal situation cannot be achieved or the phenomenon of over-fitting will occur [20].

Therefore, a novel Self-Organizing Cosine Similarity Learning Network (SO-CSLN) based on the covariance integrated the transfer entropy and the cosine similarity theory is proposed. The SO-CSLN method can be mainly used in ethylene production.

Covariance has a very important application in scientific research especially in agriculture and finance. The covariance can eliminate the influence of the different test result and improve the reliability of the experimental results. Entropy is an important measure of determining the uncertainty information based on the probability theory, sampling and communication. It can be used to solve the complex internal systems and complex diversity of the empirical distributions [24]. Chen et al. applied the entropy to the ANN and predicted the result of the rainfall-runoff simulation better [25]. Dehmer et al. demonstrated that the entropy of the graph played an important role in biology, chemistry and sociology [26]. Zou et al. introduced a new weight evaluation process using the entropy method through considering the difficulty of the fuzzy synthetic evaluation method in calculation of the multiple factors and ignorance of the relationship among evaluating objects [27]. In mathematics, cosine similarity theorems analyzed the similarity by measuring the cosine of the angle between the two vectors. It is an important metric in various fields and has a very wide range of applications [28,29]. The similarity measure is an important tool for determining the degree of similarity between two objects [30]. Lin et al. used the general cosine similarity to obtain the appearance frequency of a certain keyword in a document [31]. Nguyen et al. proposed a new method named the cosine similarity metric learning for learning a distance metric to improve the generalized ability of the acial verification [32]. Therefore, in the proposed method, the weights between the input layer and the hidden layer and between the hidden layer and the output layer are obtained by the transfer entropy and the cosine similarity theory, respectively. Meanwhile, **t**he hidden layer nodes of the SO-CSLN are determined by the rank of the sample covariance matrix based on the central limit theorem.

Comparing with the ELM, the excellent learning performance of the novel SO-CSLN model is obtained by training and demonstrating the effectiveness and stability of the proposed prediction model through UCI benchmark datasets. Further, ethylene production level is one of the main signs to determine the country industrial level. During the ethylene production process, there are many complex issues of complex production processes and the relevant factors. Meanwhile, the high cost of raw materials, excessive consumption of resources and the overcapacity issues promote us to carry out the technical reform. Therefore, a predictive modeling of the ethylene production system based on the SO-CSLN is built. The experiment results show that the novel SO-CSLN has better generalization accuracy and learning efficiency than the ELM in dealing with the complex high-dimensional and stronger coupling data. Moreover, it can guide ethylene production and improve the energy efficiency.

## 2. Extreme learning machine

The ELM can initialize the input weights and the bias and get the corresponding output weights. The structure of the ELM is shown in Fig. 1.

Supposed there are $N$random samples $(X_i, t_i)$,$X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T \in R^n$,$t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T \in R^m$. $X$is random data sample inputs, $t$ is the outputs. the single hidden layer neural network can be expressed as:

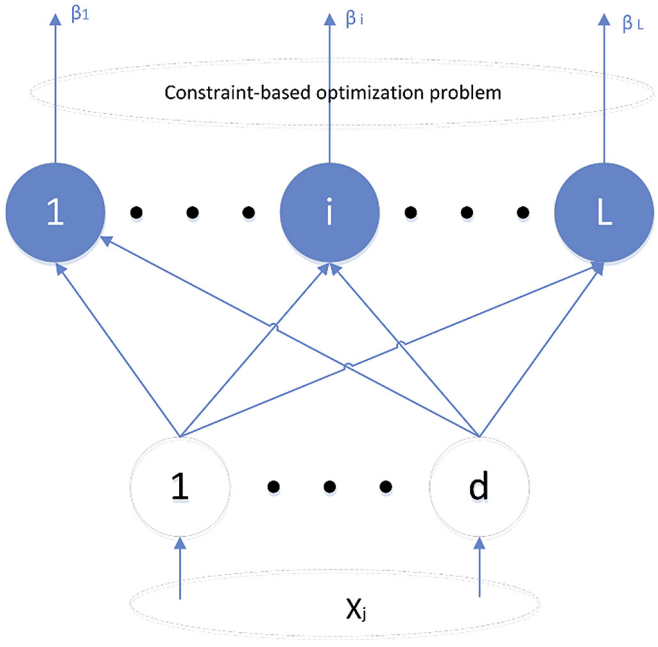


**Fig. 1.** The structure diagram of the ELM.

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

where, $L$ is the number of the hidden layer nodes, $g(x)$is the activation function,$W_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T$ is the input weights , $\beta_i$ is the output weights , $b_i$ is the bias of the $i^{th}$hidden layer units $W_i \cdot X_j$ indicates that the inner product of $W_i$ and$X_j$. The goal of the single hidden layer neural network is to obtain the minimum output error. It can be expressed as:

$$\sum_{j=1}^{N}\|o_j - t_j\| = 0 \tag{2}$$

That is the presence of$\beta_i$, $W_i$and$b_i$to make

$$\sum_{i=1}^{L}\beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, ..., N \tag{3}$$

The matrix can be expressed as:

$$H\beta = T \tag{4}$$

which in,$H$ is the output of the hidden layer nodes,$\beta$ is the output weight, $T$is the expected output.

$$H(W_1, \cdots, W_L, b_1, \cdots, b_L, X_1, \cdots, X_L) = \begin{pmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_L \cdot X_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_L \cdot X_N + b_L) \end{pmatrix}_{N*L} \tag{5}$$

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

$\widehat{W}_i$, $\widehat{b}_i$ and$\widehat{\beta}_i$are obtained as the following:

$$\left\|H\left(\widehat{W}_i, \widehat{b}_i\right)\widehat{\beta}_i - T\right\| = \min_{W,b,\beta}\left\|H(W_i, b_i)\beta_i - T\right\| \tag{6}$$

where, $i = 1, \cdots, L$. This is equivalent to minimizing the loss function:

$$E = \sum_{j=1}^{N}\left(\sum_{i=1}^{L}\beta_i g(W_i \cdot X_j + b_i) - t_j\right)^2 \tag{7}$$

In the ELM, once the $W_i$and the $b_i$are randomly determined, and the hidden layer of the output matrix$H$can be decided uniquely. The training of the single hidden layer neural network can be transformed into the linear system$H\beta = T$. And the $\beta$ can be determined as:

$$\widehat{\beta} = H^{\dagger}T \tag{8}$$

where, $H^{\dagger}$is the matrix Moore-Penrose of $H$. And it can be proved that the norm of the solution $\widehat{\beta}$ is minimal and unique.

## 3. Self-organizing cosine similarity learning network (SO-CSLN)

### 3.1. Central limit theorem and the entropy

According to the central limit theorem, $m$ samples$X = (X_1, X_2, \cdots, X_p) = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mp} \end{bmatrix}_{m*p}$ are given. $X$is data sample inputs. The sample is trend to the Gaussian distribution when the samples are infinite. All features of the sample are recognized as different Gaussian distributions to obtain the mean value vector $M = (\mu_1, \mu_2, ..., \mu_p)^T$and covariance $\sum$of the sample presented as Eqs. (9) and (10).

$$\mu_k = \frac{1}{m}\sum_{i=1}^{m}x_{ik} \tag{9}$$

$$\sum = \frac{\left((X - M^T)^T * (X - M^T)\right)}{(m-1)} \tag{10}$$

where, $p$ is the feature of the sample. Then the pivot variables are chosen as the independent characters, so the rank of the covariance matrix can be set as the node of the hidden layer, and all nodes of the hidden layer are mutual independent. Because each feature of the sample is supposed to the Gaussian distribution, the normalized input probability matrix $Z = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mp} \end{bmatrix}_{m*p}$ can be easily obtained, where $z_{ij}$ is obtained by Eq. (11).

$$z_{ij} = \left(\frac{1}{\sqrt{2*\pi}*\delta_j} * e^{-\frac{(x_{ij}-\mu_j)^2}{2*\delta_j}}\right) \Big/ \max_j \tag{11}$$

where $\delta_j = \frac{1}{m-1}\sum_{i=1}^{m}(x_{ij} - \mu_j)^2$ and $\max_j$ is the maximum value of the $jth$ feature. The primary variable is chosen by leading to the loss of information in the process of calculation. The addition of the probability value is not equal to 1. So that the $\max_j$ needed to be divided. Thus, the sum of the probability will equals 1. Meanwhile, the probability consistency can be guaranteed obviously. In order to obtain the information of the sample, the entropy of the input sample has been calculated, which is presented in Eq. (12).

$$H_j = -\sum_{i=1}^{m}z_{ij}log(z_{ij}) \tag{12}$$

where, the function $log$ is the nature logarithm. Also the total entropy of the sample is obtained by Eq. (13).

$$H = \sum_{j=1}^{p}H_j \tag{13}$$

The information of the input sample will flow into every node in the hidden layer, and the information flow between the input layer and the hidden layer, which is seen as the weight $\gamma = \begin{bmatrix} \gamma_1 & \cdots, & \gamma_p \end{bmatrix}^T = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1Q} \\ \vdots & \ddots & \vdots \\ \gamma_{p1} & \cdots & \gamma_{pQ} \end{bmatrix}_{p*Q}$ between the input layer and the hidden layer. And weights are defined by Eq. (14). Where $Q$ is the hidden nodes.

$$Y_{ij} = e^{-\frac{|H_i - H*rdmID(j)|}{H*rdmID(j)}} \tag{14}$$

where, the $rdmID$will generate the non-redundant data randomly as shown in Eq. (15).

$$\sum_{i=1}^{Q} rdmID(i) = 1 \tag{15}$$

### 3.2. Cosine similarity learning network (CSLN)

Given $m$ samples $X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mp} \end{bmatrix}_{m*p}$ and $Y = [y_1 \quad ,\ldots, \quad y_s] = \begin{bmatrix} y_{11} & \cdots & y_{1s} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{ms} \end{bmatrix}_{m*s}$ respectively. Where $p$ is input features and $s$ are outputs. And there are $Q$ hidden nodes in SLFNs. The outputs in the hidden layer of all samples are denoted by $H = \begin{bmatrix} h_{11} & \cdots & h_{1Q} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mQ} \end{bmatrix}_{m*Q}$.

The weight $\gamma_{p+1} = [\gamma_{(p+1)1} \quad ,\ldots, \quad \gamma_{(p+1)Q}]$ between the input layer and the hidden layer are generated randomly, which are restricted to [-1, 1]. Considering the bias of all hidden nodes, an extend node has been put in the input layer whose bias is 1. Therefore, the extend samples $Xext = \begin{bmatrix} x_{11} & \cdots & x_{1p} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mp} & 1 \end{bmatrix}_{m*(p+1)}$ are got, and the outputs of the hidden layer are obtained by Eq. (16).

$$H = Xext*\gamma \tag{16}$$

The network will produce a group of outputs denoted by $O = [o_1 \quad ,\cdots, \quad o_s] = \begin{bmatrix} o_{11} & \cdots & o_{1s} \\ \vdots & \ddots & \vdots \\ o_{m1} & \cdots & o_{ms} \end{bmatrix}_{m*s}$ when a group of training samples are given. And the outputs of the output layer are obtained by Eq. (17).

$$O = Hext*W \tag{17}$$

where $W$is the weight between the hidden layer and the output layer denoted by$W = [w_1 \quad ,\ldots, \quad w_s] = \begin{bmatrix} w_{11} & \cdots & w_{1s} \\ \vdots & \ddots & \vdots \\ w_{Q1} & \cdots & w_{Qs} \end{bmatrix}_{Q*s}$, and $Hext$ is an extend matrix of $H$ obtained by Eq. (18).

$$Hext = H + [E \quad Z]^T \tag{18}$$

where $E$ is an identity matrix of $Q$ by $Q$, and $Z$ is a zero matrix of $Q$ by $m$-$Q$.

The matrixes $Y$ and $O$ can be seen as some separate column vectors, which denoted by $C(Y)$and $C(O)$, respectively, so that the $C(Y) = \{y_1, y_2, \ldots, y_s\}$and the $C(O) = \{o_1, o_2, \ldots, o_s\}$can be got. Meanwhile, the cosine included angle formula between column vectors $C(Y)$ and $C(O)$ is defined, respectively, which are presented in Eq. (19). $J(W)$ is the function of similarity computation.

$$J(W) = Cosine\langle C(Y), C(O)\rangle = \prod_{i=1}^{s} Cosine\langle C(Y)_i, C(O)_i\rangle \tag{19}$$

where $C(Y)_i$shows the $ith$ column vector in $C(Y)$and$C(O)_i$ shows the $ith$ column vector in $C(O)$. Then the weight $W$ can be obtained according to Eq. (20) because we the included angle between the column vectors of $Y$ and the column vectors of $O$ need to be minimized and ensure that the SO-CSLN can obtain correct answers given samples $[X\ Y]$. $W$is the minimum value in similarity.

$$W = \underset{W}{Argmin}\ J(W) \tag{20}$$

The detail derivational process is listed in Appendix A, which can solve Eq. (20).

By solving Eq. (20), the weight $W$ presented at Eq. (21) can be obtained.

$$W = \left(Hext^T*Hext\right)^{-1}*Hext^T*Y \tag{21}$$

Finally, the outputs of the network can be got by Eq. (22).

$$O = H*W \tag{22}$$

### 3.3. The flow of the SO-CSLN algorithm

Step 1: Select the training and testing data sets.

Step 2: Preprocess and normalize the data of the datasets by Eq. (23)

$$\overline{x}_j^t = \begin{cases} \dfrac{x_j^{(t)} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}, if x_j^{\max} \neq x_j^{\min} \\ -1, other \end{cases} \tag{23}$$

where in, $x_j^{(t)}$ indicates the variable value of the sample,$x_j^{\max} = \max\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$,$x_j^{\min} = \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$, $i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$.

Step 3: The mean value vector and the covariance can be got by Eq. (9) and Eq. (10). The numbers of the hidden layer nodes are determined by the rank of the covariance. The normalized input probability matrix is obtained by Eq. (11).

Step 4: The total information entropy can be achieved by Eq. (12) and Eq. (13). The weights between the input layer and the hidden layer are determined by Eq. (14) and Eq. (15). The $rmdID$will generate the non-redundant data randomly.

Step 5: The output of the hidden layer is obtained by Eq. (16).

Step 6: The included angle between the two column vectors is minimized through Eq. (19). And obtain the weights between the hidden layer and the output layer by Eq. (20) and Eq. (21).

Step 7: After the anti-normalization process, the final prediction results are obtained by Eq. (24).

$$x_j^{(t)} = \overline{x}_j^{(t)} * \left(X_j^{\max} - X_j^{\min}\right) + X_j^{\min} \tag{24}$$

Step 8: Get the output result by Eq. (22).

The structure of the novel SO-CSLN is shown in Fig. 2.

## 4. Benchmark test of the UCI data sets

### 4.1. Benchmark test of the regression data sets

In order to further validate the efficiency of our proposed algorithm, the Housing, Airfoil and abalone data sets from the UCI repository are chosen [33]. The average relative generalization error (ARGE) and the root mean square error (RMSE) are obtained by Eq. (25) and Eq. (26).

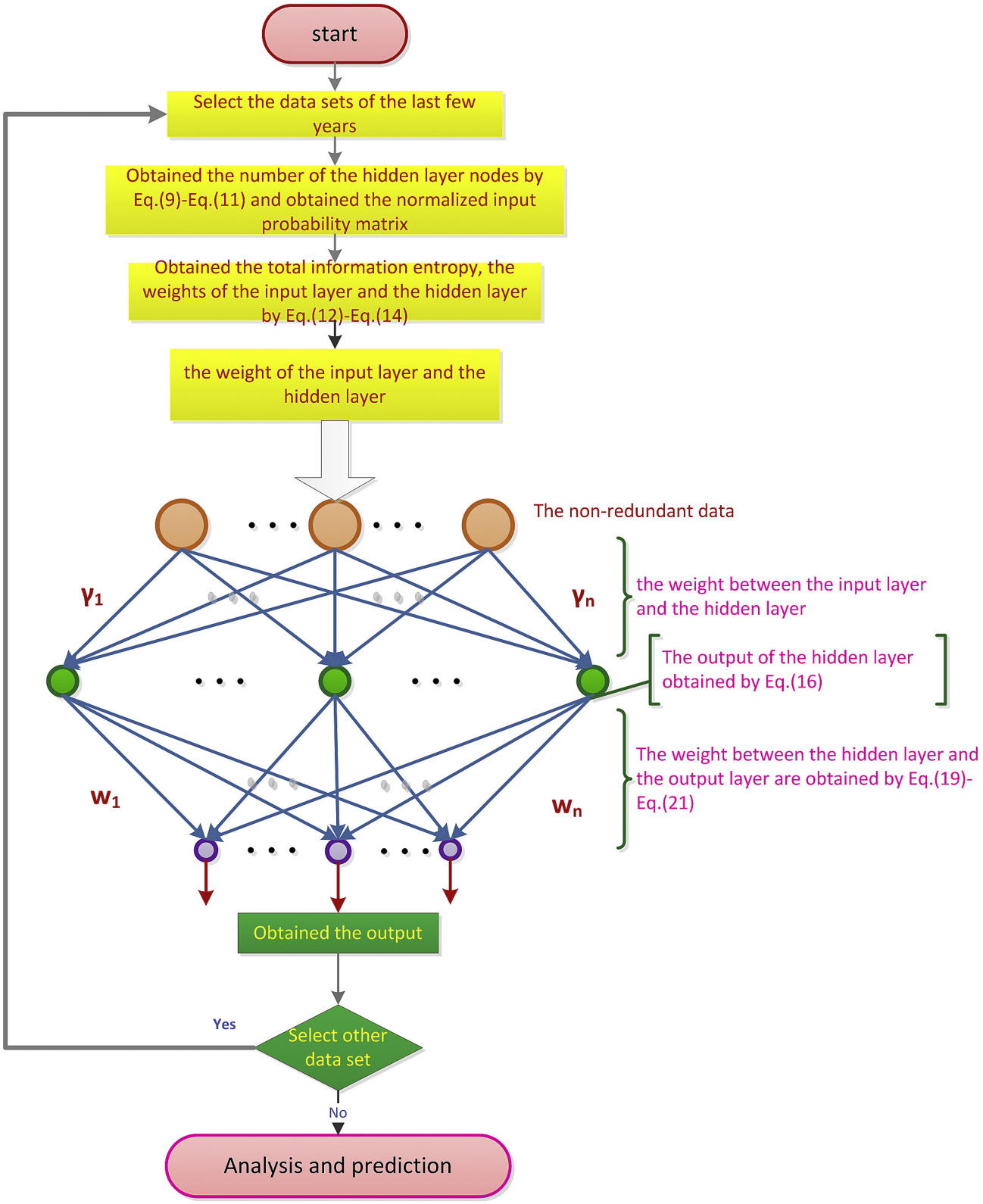


**Fig. 2.** The flow diagram of the novel SO-CSLN.

**Table 1**
Specification of the UCI data sets.

| Data Sets | #Samples | | #Attributes | |
|---|---|---|---|---|
| | Training | Testing | Inputs | Outputs |
| Housing | 338 | 168 | 13 | 1 |
| Airfoil | 1000 | 503 | 5 | 1 |
| Abalone | 2000 | 2177 | 7 | 1 |

$$\begin{cases} MSE_i = \left(NetOut_i^{norm} - ExpectOut_i^{norm}\right)^2 \\ RMSE = \sqrt{\dfrac{\sum_{i=1}^{n} MSE_i}{n}} \end{cases} \tag{26}$$

As is shown in Table 1, the regression analysis dataset of housing, airfoil and abalone has been used to verify the suitability and robustness of the SO-CSLN. The SO-CSLN is a self-organizing neural network, so the experimenter does not need to set the hidden layer nodes. The program determines the number of the hidden layer nodes based on the samples. So the'N' represents there is no corresponding hidden layer nodes so that there is no training results.

When analyzing the housing, airfoil and abalone data sets, the program automatically determine that the hidden layer nodes are 13, 5 and 7 based on the sample.

As shown in Table 2 and Fig. 3, in the housing, airfoil and abalone

**Table 2**
Comparison of the RMSE and the ARGE.

| Data set | Node | SO-CSLN | | | | ELM | | | | BP | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Train | | Test | | Train | | Test | | Train | | Test | |
| | | RMSE | ARGE (%) | RMSE | ARGE (%) | RMSE | ARGE (%) | RMSE | ARGE (%) | RMSE | ARGE (%) | RMSE | ARGE (%) |
| Housing | 9 | N | N | N | N | 0.21 | 37.33 | 0.216 | 29.56 | 0.251 | 22.545 | 0.274 | 24.749 |
| | 13 | 0.10 | 17.6 | 0.095 | 17.19 | 0.194 | 39.66 | 0.196 | 26.78 | 0.203 | 19.339 | 0.206 | 19.783 |
| | 15 | N | N | N | N | 0.145 | 24.08 | 0.155 | 20.41 | 0.212 | 21.199 | 0.222 | 24.161 |
| Airfoil | 3 | N | N | N | N | 0.364 | 4.461 | 0.404 | 4.988 | 0.347 | 4.813 | 0.400 | 5.336 |
| | 5 | 0.28 | 3.32 | 0.34 | 4.053 | 0.299 | 3.681 | 0.409 | 4.977 | 0.303 | 3.984 | 0.421 | 5.206 |
| | 7 | N | N | N | N | 0.281 | 3.345 | 0.422 | 4.890 | 0.299 | 3.312 | 0.305 | 6.389 |
| Abalone | 5 | N | N | N | N | 0.194 | 19.70 | 0.226 | 19.99 | 0.184 | 19.51 | 0.175 | 19.121 |
| | 7 | 0.19 | 19.1 | 0.22 | 18.75 | 0.181 | 19.34 | 0.225 | 21.29 | 0.172 | 20.099 | 0.178 | 21.686 |
| | 9 | N | N | N | N | 0.164 | 17.19 | 0.226 | 24.12 | 0.181 | 18.934 | 0.185 | 18.800 |

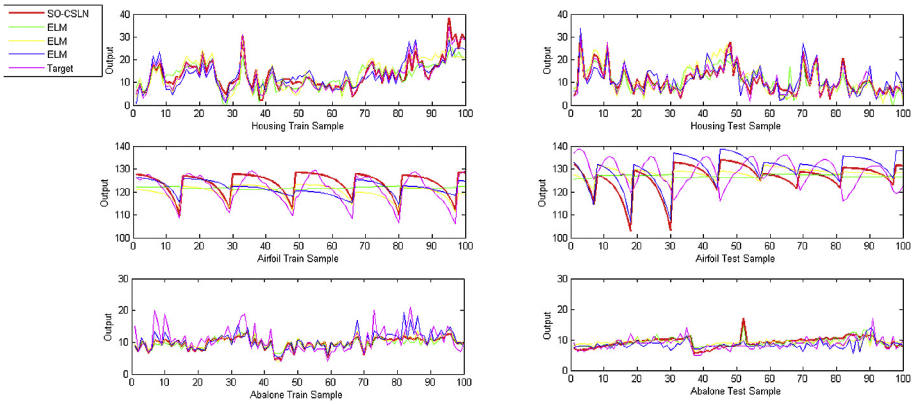


**Fig. 3.** RMSE and ARGE of the ELM and the SO-CSLN.

$$\begin{cases} \begin{cases} RGE_i = Abs\left(\dfrac{NetOut_i^{inver} - ExpectOut_i^{inver}}{Expect_i^{inver}}\right), ifExpectOut_i^{inver} \neq 0 \\ RGE_i = Abs\left(NetOut_i^{inver} - ExpectOut_i^{inver}\right), else \end{cases} \\ ARGE = \dfrac{\sum_{i=1}^{n} RGE_i}{n} * 100 \end{cases} \tag{25}$$

**Table 3**
Specification of the UCI data sets.

| Data Sets | #Samples | | #Attributes | |
|---|---|---|---|---|
| | Training | Testing | Inputs | Outputs |
| CMC | 1000 | 473 | 9 | 1 |
| Iris | 144 | 6 | 4 | 1 |

**Table 4**
Classification accuracy test of the SO-CSLN and the ELM.

| Data Set | Node | SO-CSLN | | ELM | | BP | |
|---|---|---|---|---|---|---|---|
| | | Train | Test | Train | Test | Train | Test |
| Iris | 2 | N | N | 0.513 | 0.500 | 0.655 | 0.612 |
| | 4 | 0.979 | 0.999 | 0.847 | 0.666 | 0.778 | 0.680 |
| | 6 | N | N | 0.937 | 0.8333 | 0.863 | 0.857 |
| CMC | 6 | N | N | 0.417 | 0.441 | 0.403 | 0.415 |
| | 9 | 0.490 | 0.505 | 0.454 | 0.469 | 0.434 | 0.422 |
| | 12 | N | N | 0.498 | 0.486 | 0.458 | 0.427 |

data sets, the RMSE, the ARGE of the SO-CSLN are better than that of the ELM and the BP. Huang has proved that the RMSE, the ARGE and the success rate of the ELM is better than the BP in most cases [17]. So in this paper, the ELM is mainly used to compare with the SO-CSLN. In table two, the RMSE of the airfoil test set is 0.342 and the ARGE is 4.053%. The RMSE and the ARGE of the ELM is 0.409 and 4.977% respectively with the same number of nodes in the hidden layer. The RMSE and the ARGE of the BP are 0.421 and 5.206%. In general, the more hidden layer nodes, the smaller the error will be. It can be seen from Fig. 3, in housing, airfoil and abalone data sets, the hidden layer nodes of the ELM are 15, 7 and 9, respectively. And compared with the hidden layer nodes of the SO-CSLN with 13, 5 and 7, the RMSE and the ARE of the ELM are larger. Meanwhile, Table 2 illustrates the SO-CSLN has a higher generalization precision, a smaller error and a better stability compared with the ELM.

### 4.2. Benchmark test of the classical data sets

In order to verify the classification accuracy of the SO-CSLN, we use iris and CMC [33] from the UCI repository in Table 3. The classification accuracy is obtained by Eq. (27) [34].

$$Classification\ \ Accuracy = \frac{Correct\ \ Samples}{Total\ \ Samples} \tag{27}$$

Table 4 shows the contrast of the SO-CSLN and the ELM classification accuracy. When testing the iris and CMC data sets, the SO-CSLN program automatically sets the hidden layers are 4 and 9 based on the sample.

Based on the iris and the CMC data sets, the accuracy of the classification accuracy of the SO-CSLN, the ELM and the BP is shown in Table 4. When the SO-CSLN, the ELM and the BP have same hidden layer nodes, the correct rate of the SO-CSLN is higher. For example, the classification accuracy of the iris test set is 0.999, the ELM is 0.666 and the BP is 0.680 with the same number of nodes in the hidden layer. Meanwhile, Table 4 illustrates the SO-CSLN has better effectiveness and stability.

## 5. Ethylene production prediction modeling

### 5.1. Ethylene input and output data analysis

Cracking and separation are two main parts of ethylene production [35]. When the cracking furnace is running, the cracking reaction tube is performing. Burning a lot of fuel to provide heat and TLE (transmission line) produces a large number of steams by recovering the waste heat. In order to make the raw material hydrocarbon completing the best cleavage reactions in a short period of time to reduce coke, the steam should be poured into the cracking furnace while the hydrocarbon is put into it. Separation portion mainly comprises a rapid cooling, compression and a separate section. The power of the ethylene equipment has three scopes: 200 thousand tons, 200–600 thousand tons and more than 600 thousand tons. The main energy consumption is the power consumption of the compressor, the heat separating steam and cooling energy consumption of the compressor and the cold box [36]. The main energy consumption in cracking section includes the preheat of the mixture of the raw materials and the stream, the reaction heat consumption in the cracking reactions, and the waste heat released to the environment like afterheat in the flue gas [34]. A typical framework of the ethylene plant flowchart is shown in Fig. 4.

In order to verify the feasibility and the effectiveness of the novel SO-CSLN in the real world, this paper selects the monthly production data of 19 ethylene plants from 2009 to 2013 under 7 major ethylene production technologies in China as analysis object. In the ethylene industry, different companies use different energy efficiency sector analysis of ethylene systems. In order to analyze the ethylene production and the energy efficiency of ethylene plants better, dividing ethylene plant precinct referenced into ethylene industry data standard based on DB 37/751-2007 and GB/T 2589-2008 [37,38]. The factors associated with ethylene production efficiency are: raw materials, fuel, power consumption and products. Meanwhile, according to the relation of Chemical Engineer Design for Energy Consumption Calculation Method conversion [39], the unit measurement of consumption related parameters is converted such as fuel, steam, water, and electricity into GJ. The unit of the crude oil and outputs are tons. The five categories of raw materials have been put as the inputs of the network. Finally, 5 inputs and 1 output training data are constituted.

The crude oil includes naphtha, light diesel oil, raffinate oil, hydrocracking bottom oil 1, hydrocracking bottom 2 and carbon 345 (carbon 345 means the sum of carbon 3, carbon 4 and carbon 5). The fuel includes light oil, heavy oil and fuel gas. The steam

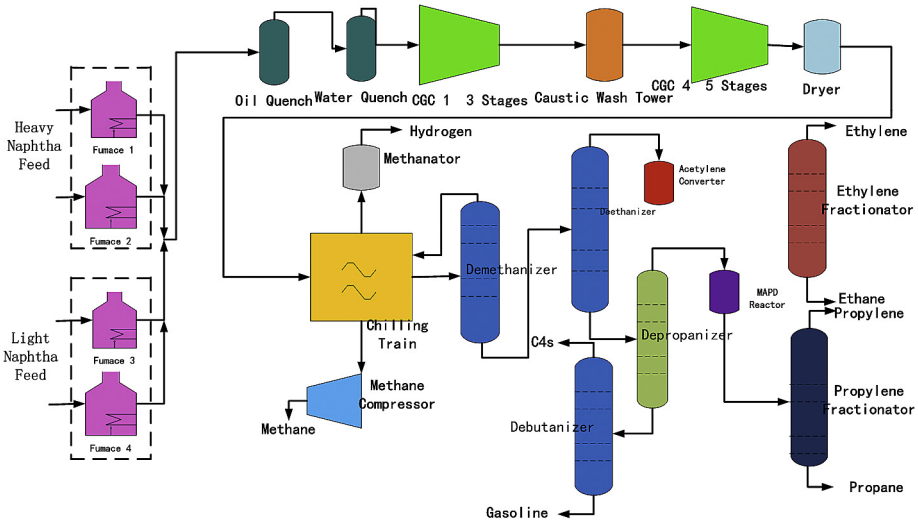


**Fig. 4.** Ethlene production plant diagram.

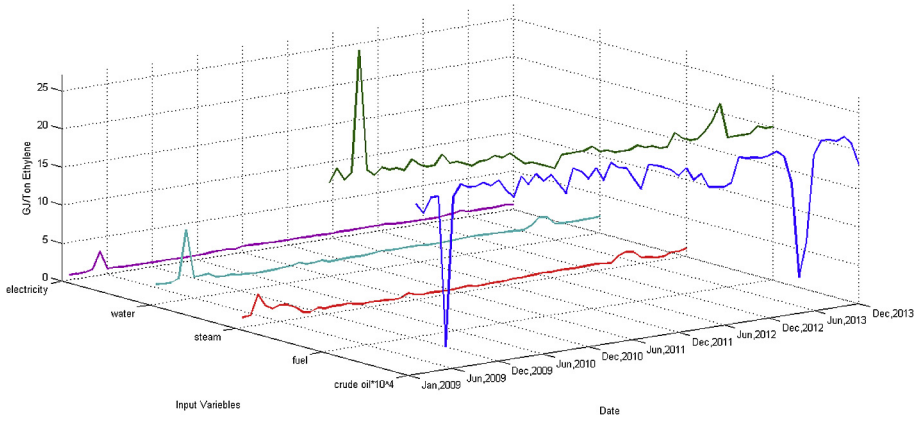


**Fig. 5.** Energy consumption data of the plant 1 in 2009–2013.

**Table 5**
The contrast of the ARGE under the same scale and different technologies.

| SO-CSLN | ELM |
|---|---|
| 2.860% | 4.675% |

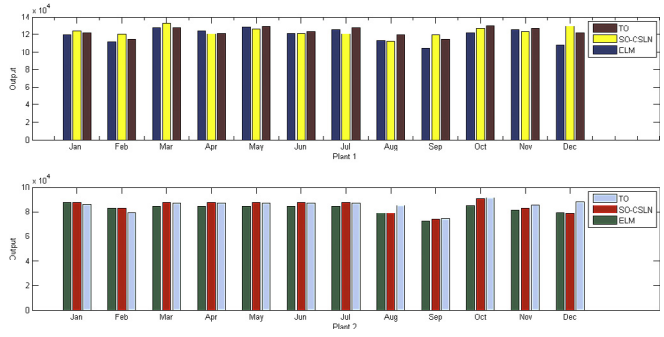


**Fig. 6.** Prediction results of ethylene production plants with the same scale and different technologies.

The crude oil, fuel, steam, water and electricity and the sun of ethylene, propylene and carbon 4 are chosen as the input data and the output, respectively.

### 5.2. Production prediction and comparison of different Ethylene plants

#### 5.2.1. Same scale and different technologies

In order to improve the ethylene production, choosing the real energy data of some ethylene plants under the same scale and different technologies in 2009 to build the prediction model to predict the ethylene production in 2010. In the scale of 800 thousand tons of plants, there are three kinds of technologies consisted of Lummus, S&W and Linde. The ELM has been used as a contrast. The ARGE of the SO-CSLN and the ELM is shown in Table 5 by using ethylene data sets. The ARGE of the SO-CSLN and the ELM is 2.860% and 4.675%, respectively. The prediction accuracy of the SO-CSLN is 1.63 times than the ELM. The smaller error in ethylene industrial data, the better prediction accuracy of the industrial production will be got. So the production of the ethylene will increase by 1.815%.

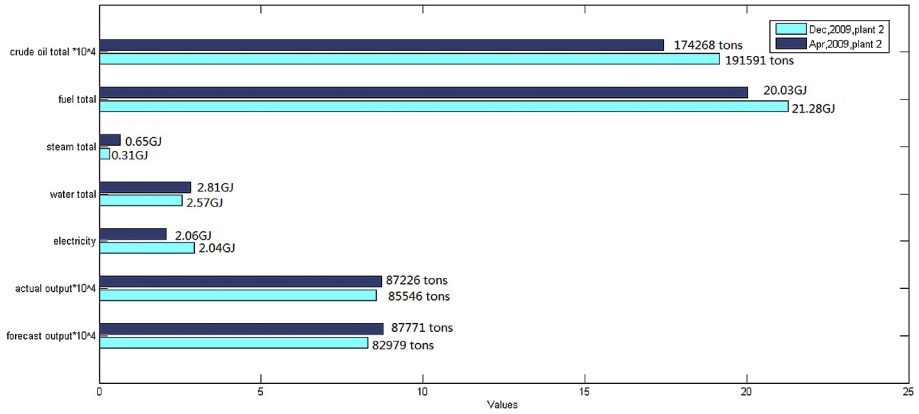


**Fig. 7.** Production contrast diagram of the plant 2.

includes ultra-high pressure steam, high pressure steam, medium pressure steam and low pressure steam. The water includes circulating water, industrial water, boiler water and other water. The data is added up directly. The input data of Plant 1 is shown in Fig. 5.

Comparing with the ELM, we use two different data sets (Plant 1 and Plant 2) to analysis ethylene production and energy efficiency. The result of target output (TO), SO-CSLN and ELM is shown in Fig. 6.

As is shown in Fig. 6, the ARGE of ethylene production of the plant 2 is less than that of the plant 1. It shows that the plant 2 is more stable and the production is satisfactory. The production volatility of the plant 1 is larger, indicating that the energy efficiency is lower. Therefore, according to the production situation, it can be adjusted the input raw materials and the ethylene production of the plant 2 to achieve the stable production and improve the energy efficiency as shown in Fig. 7.

It can be seen in Fig. 7 that the plant 2 put into 174286 tons crude oils, 20.03 GJ fuel, 0.65 GJ steams, 2.81 GJ water, 2.06 GJ electricity for production and the total actual output is 87226 tons according with the prediction output with 87771 tons complied with the requirement for full load operation. The plant 2 can maintain the production status in subsequent production. The plant 2 invests 191591 tons crude oils, 21.28 GJ fuel, 0.31 GJ steams, 2.57 GJ water, 2.04 GJ electricity and the actual output is 85546 tons far larger than the predicted model output (82979 tons). it results in the overload operation. To further improve the production efficiency, the input of raw materials in April can be referred and the energy consumption is reduced to improve energy efficiency. The production can be increased to 1.964%.

As is shown in the Fig. 7, the crude oil investment in April was 17323 tons less than December. The Fuel investment in April is 1.23 GJ less than December. The electricity in April was 0.02 GJ than December. So the inputs of crude oil, fuel and electricity in December have exceeded saturation. Especially the crude oil and the fuel waste too much. Therefore, the amount inputs of these raw materials need to be reduced. In April, the input of the steam and the water are 0.34 GJ and 0.24 GJ more than December. It should be considered increasing the amount of the water vapor in production.

Therefore, the plant 1 can refer to the production analysis of the plant 2 of full load, over load and insufficient load to improve the production situation and achieve full load operation.

### 5.2.2. *Same technology and different scales*

At the same time, selecting two ethylene plants of 200 thousand tons and 800 thousand tons under the S&W technology in 2009 and 2010 to analyze and predict the production situation. And then the result is shown in Table 6. The ARGE of the SO-CSLN is 2.158% and the ELM is 5.172%. The prediction accuracy of the SO-CSLN is 2.39 times than the ELM. So the production of the ethylene will increased by 3.014%. The prediction of the novel SO-CSLN is better than the ELM under the two plants.

The result of target output (TO), SO-CSLN and ELM of the plant 3 and the plant 4 is shown in Fig. 8. Meanwhile, ethylene production and energy efficiency analysis of the plant 4 is shown in Fig. 9.

As is shown in Figs. 8 and 9 that the plant 4 invests crude oils 50456 tons, fuel 17.94 GJ, steams 5.7 GJ, water 2.06 GJ, electricity 1.02 GJ and the total actual output is 29287 tons based on the prediction output with 30208 tons. The plant 4 is full load operation in February and can maintain the production status in subsequent production. The plant 4 invests crude oils 48538 tons, fuel 18.36 GJ, steams 2.19 GJ, water 6.63 GJ and electricity 1.19 GJ in June. The actual output is 28242 tons and it is larger than the predicted output (30421 tons). It shows that the plant 4 is insufficient load production in June. In order to further improve the energy efficiency, referring to the input of raw material in February to reduce the energy consumption and improve the ethylene production efficiency. The production in June can increase about 0.705%.

**Table 6**
The contrast of the ARGE under the same technology and different scales.

| SO-CSLN | ELM |
|---|---|
| 2.158% | 5.172% |

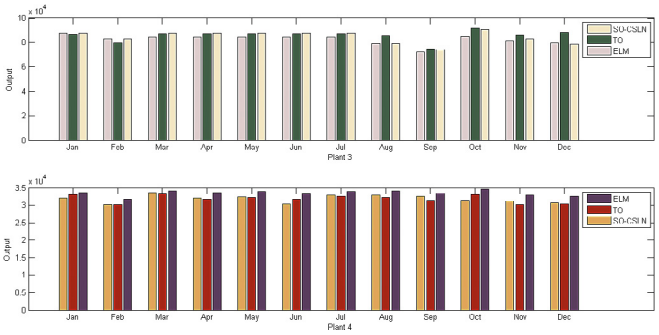


**Fig. 8.** Prediction results of ethylene production plants with the same technology and different scales.

As shown in the Fig. 9, the output in February is 1045 tons more than in June. The most significant difference between February and June is the consumption of fuel. Therefore, the consumption of fuel input can be increased appropriately.

High yields of ethylene plants have a larger amount of investment, and there are many influencing factors in the production process. According to the prediction model, we can adjust the input of each month to reach the best output.

## 6. Discussion

First, the novel SO-CSLN is proposed. And then the validity and robustness of the novel SO-CSLN are verified effectively by UCI standard data sets. Meanwhile, compared with the ELM, the novel SO-CSLN has a smaller error and more accurate prediction.

Second, the production of the same scale of different technologies and the same technology of different scales in ethylene production processes are analyzed and predicted by using the novel SO-CSLN. Based on the experiment result, the input of the crude oil, fuel, steam, water and electricity can be made in the more reasonable proportion to improve energy efficiency.

Third, this proposed model is self-organized and the number of hidden layer nodes is determined by the rank of covariance based on the input samples. Therefore, we will add the self-adaptive method to this proposed model and make it more suitable for the production of the real world.

## 7. Conclusion

In this paper, a novel SO-CSLN based on the central limit theorem, the information entropy and the cosine theorem is proposed. The number of the hidden layer nodes can be obtained by the rank of the input sample covariance matrix. The weights between the input layer and the hidden layer and between the hidden layer and the output layer are obtained by the entropy and the cosine similarity theory, respectively. Meanwhile, the stable structure can be obtained based on the central limit theorem and the SO-CSLN method can be validated through UCI benchmark datasets and ethylene production systems. Compared with the ELM, the novel SO-CSLN has faster convergence rate, higher learning accuracy and stronger generalization. Simultaneously, the SO-CSLN could predict the ethylene production of different technologies and different scales accurately based on the input and output configuration of the ethylene production. Meanwhile, it can increase the ethylene production to 1.964% and 0.705%, respectively. Moreover, the SO-CSLN has an important theoretical and practical value for other complex

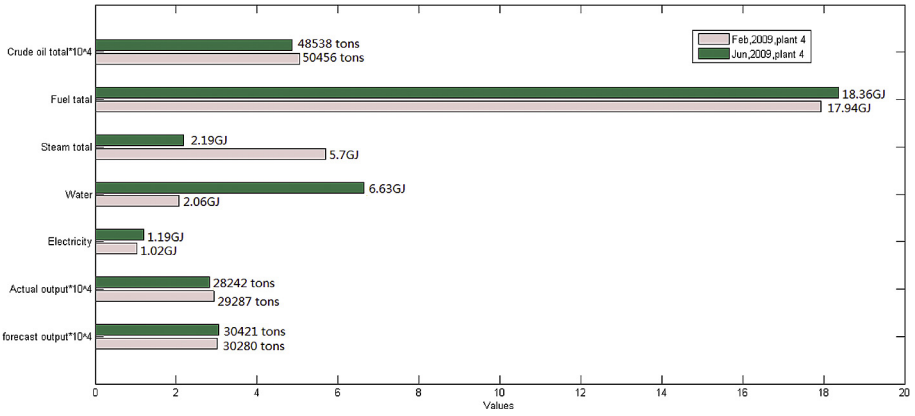


**Fig. 9.** Production contrast diagram of the plant 4.

industrial processes.

In our further studies, we will explore and integrate other approaches, such as self-adaptive method and particle swarm optimization (PSO). Moreover, considering the economic, environmental and human factors on the experiment results and the weight of different factors in the production prediction of complex chemical industries, we will refine the input data to further improve the prediction accuracy.

## Acknowledgments

This research was partly funded by National Natural Science Foundation of China (61673046, 61533003, 61603025), Natural Science Foundation of Beijing (4162045) and the Fundamental Research Funds for the Central Universities(ZY1703, JD1708).

## Symbols used

| | |
|---|---|
| SLFN | single layer feed-forward network |
| SO-CSLN | self-organizing cosine similarity learning network |
| MGD | microgrid distribution |
| MG | microgrid |
| PV | photo voltaic |
| FC | fuel cell |
| ANN | artificial neural network |
| DSS | decision support system |
| BP | back propagation |
| RBF | radial basis function |
| ENN | Elman neural network |
| MPPT | maximum power point tracking |
| ELM | extreme learning machine |
| CSLN | cosine similarity learning network |
| ARGE | average relative generalization error |
| RMSE | root mean square error |
| TLE | transmission line |
| carbon 345 | carbon 345 means the sum of carbon 3, carbon 4 and carbon 5 |
| TO | target output |
| PSO | particle swarm optimization |
| FLNRFNN | functional link-based novel recurrent fuzzy neural network |

## Appendix A

We have $J(W) = \langle CosineC(Y), C(O) = \prod_{i=1}^{s} CosineC(Y)_i, C(O)_i \rangle$, so we can get likelihood of $J(W)$ listed at Eq. (A-1).

$$L(W) = \log J(W) = \sum_{i=1}^{s} Cosine\langle C(Y)_i, C(O)_i \rangle \tag{A-1}$$

Then minimizing $J(W)$ is the same as minimizing $L(W)$, so we get Eq. (A-2).

$$W = \underset{w}{Argmin}\, J(W) = \underset{w}{Argmin}\, L(W) \tag{A-2}$$

Therefore, we should minimize each item of $L(W)$, so Eq. (A-2) can be reduced to Eq. (A-3).

$$w_i = \underset{w_i}{Argmin}\ J(w_i) = \underset{w_i}{Argmin}\ L(w_i) \tag{A-3}$$

and

$$L(w_i) = ine\langle C(Y)_i, C(O)_i \rangle = \frac{Y_i^T * Hext * w_i}{\|Y_i\| * \|o_i\|} \quad \{\|Y_i\| * \|o_i\| \neq 0\} \tag{A-4}$$

In order to minimize the$L(w_i)$, so we have Eq. (A-5).

$\frac{\partial L}{\partial w_i} = 0 \downarrow$

$$\overset{w_i}{\Rightarrow} Hext^T * y_i * w_i^T * Hext^T * Hext * w_i = Hext^T * Hext * w_i * y_i^T * Hext * w_i \tag{A-5}$$

And we define$A_i = Hext^T * y_i$, $B = Hext^T * Hext$, so we get Eq. (A-6).

$$A_i * w_i^T * B * w_i = B * w_i * A_i^T * w_i \quad \{w_i \neq 0\} \tag{A-6}$$

Firstly, we see $B * w_i$ as a constant, and then we can obtain Eq. (A-7) by Eq. (A-6).

$$B * w_i = A_i \Rightarrow w_i = B^{-1} * A_i \Rightarrow w_i = \left(Hext^T * Hext\right)^{-1} * Hext^T * y_i \tag{A-7}$$

Therefore, we can obtain the weight W by Eq. (A-8).

$$W = [\, w_1 \quad \ldots \quad w_i \quad \ldots \quad w_s \,] = \left(Hext^T * Hext\right)^{-1} * Hext^T * Y \tag{A-8}$$

## References


[1] Ou Ting-Chia, Su Wei-Fu, Liu Xian-Zong, et al. A modified bird-mating optimization with hill-climbing for connection decision of transformers. Energies 2016;9(9):671.
[2] Ou Ting-Chia, Chuang Shang-Jen, Hong Chih-Ming, et al. Self-regulation ground faults model for microgrid distribution. ICIC Express Lett 2015;6(12). pp.1-ICICIC.
[3] Ou Ting-Chia, et al. A novel unsymmetrical faults analysis for microgrid. IJEPES 2012;43(1):1017–24.
[4] Ou Ting-Chia, Hong Chih-Ming. Dynamic operation and control of microgrid hybrid power systems. Energy 2014;66:314–23.
[5] Han YongMing, Geng ZhiQiang, Zhu QunXiong. Energy optimization and prediction of complex petrochemical industries using an improved artificial neural network approach integrating data envelopment analysis. Energy Convers Manag 2016;124:73–83.
[6] Ou TC, Lu KH, Huang CJ. Improvement of transient stability in a hybrid power multi-system using a designed NIDC (novel intelligent damping controller). Energies 2017;10(4):488.
[7] Thakur GS, Bhattacharyya R, Mondal SS. Artificial neural network based model for forecasting of inflation in India. Fuzzy Inf Eng 2016;8(1):87–100.
[8] Khan IYPZ, Suralkar SR. Importance of artificial neural network in medical diagnosis disease like acute nephritis disease and heart disease. Int J Eng Sci Innov Technol 2013;2(2):7.
[9] Chen T, Wang YC. Estimating simulation workload in cloud manufacturing using a classifying artificial neural network ensemble approach. Robotics Comput.-Integrated Manuf 2016;38:42–51.
[10] Bal Beşikçi E, Arslan O, Turan O, et al. An artificial neural network based decision support system for energy efficient ship operations. Comput Oper. Res 2015;66(C):393–401.
[11] David E, Rumelhart, Hinton Geoffrey E, Williams Ronald J. Learning representations by back-propagating errors. Nature 1986;323:533–6.
[12] Song EZ, Liu JG, Ding SL, et al. Study of neural network control algorithm in the diesel engine. In: SAE 2016 commercial vehicle engineering congress; 2016.
[13] Kuo RJ, Shieh MC, Zhang JW, et al. The application of an artificial immune system-based back-propagation neural network with feature selection to an RFID positioning system. Robotics Comput.-Integrated Manuf 2013;29(6):431–8.
[14] Ahmadi SH, Sepaskhah AR, Andersen MN, et al. Modeling root length density of field grown potatoes under different irrigation strategies and soil textures using artificial neural networks. Field Crops Res 2014;162(6):99–107.
[15] Deosarkar MP, Sathe VS. Predicting effective viscosity of magnetite ore slurries by using artificial neural network. Powder Technol 2012;219(3):264–70.
[16] Powell MJD. The theory of radial basis function approximation in 1990. 1990.
[17] Hong Chic-Ming, Ou Ting-Chic, Lu Kai-Hung. Development of intelligent MPPT (maximum power point tracking) control for a grid-connected hybrid power generation system. Energy 2012;50(1):270–9.
[18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. Proc int joint Conf neural Netw 2004;2:985–90.
[19] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1–3):489–501.
[20] Aghbashlo M, Shamshirband S, Tabatabaei M, et al. 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.
[21] Li S, Goel L, Wang P. An ensemble approach for short-term load forecasting by extreme learning machine. Appl Energy 2016;170:22–9.
[22] Kariminia S, Shamshirband S, Motamedi S, et al. A systematic extreme learning machine approach to analyze visitors' thermal comfort at a public urban space. Renew Sustain Energy Rev 2016;58:751–60.
[23] Wang S, Deng C, Lin W, et al. A novel NMF-based image quality assessment metric using extreme learning machine. IEEE Trans Cybern 2016:255–8.
[24] Rajaram R, Castellani B. An entropy based measure for comparing distributions of complexity. Phys A-stat. Mech Its Appl 2016:35–43.
[25] Chen L, Singh VP, Guo S, et al. Copula entropy coupled with artificial neural network for rainfall–runoff simulation. Stoch Environ Res Risk Assess 2014;28(7):1755–67.
[26] Dehmer M, Mowshowitz A. A history of graph entropy measures. Inf Sci 2011;181(1):57–78.
[27] Zou ZH, Yun Y, Sun JN. Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. J Environ Sci 2006;18(5):1020–3.
[28] Al-Anzi FS, Abuzeina D. Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing. J King Saud Univ - Comput Inform Sci 2017;29(2):189–95.
[29] Xia P, Zhang L, Li F. Learning similarity with cosine similarity ensemble. Inf Sci 2015;307:39–52.
[30] Hou XN, Ding SH, Ma LZ, et al. Similarity metric learning for face verification using sigmoid decision function. Vis Comput 2015:1–12.
[31] Lin ZC, Wu DW, Hong GE. Combination of improved cosine similarity and patent attribution probability method to judge the attribution of related patents of hydrolysis substrate fabrication process. Adv Eng Inf 2016;30(1):26–38.
[32] Nguyen HV, Bai L. Cosine similarity metric learning for face verification"// computer vision - ACCV 2010. In: Asian conference on computer vision, queenstown, New Zealand, november 8-12, 2010, revised selected papers; 2010. p. 709–20.
[33] http://archive.ics.uci.edu/ml/datasets.
[34] Han Y, Geng Z, Zhu Q, et al. Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry. Energy 2015;83:685–95.
[35] 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(11–12):1279–84.
[36] Han Yongming, Geng Zhiqiang. Energy efficiency hierarchy evaluation based on data envelopment analysis and its application in a petrochemical process. Chem Eng Technol 2014;37(12):2085–95.
[37] DB37/751-2007, Limitation of energy consumption for Ethylene Product.
[38] GB/T2589-2008, General computing guide of Special Energy Consumption.
[39] SH/T3110-2001. Calculation method for energy consumption in petrochemical engineering Design. 2002.