Applied Thermal Engineering 160 (2019) 114072

Contents lists available at ScienceDirect

Applied Thermal Engineering

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

Research Paper

Production capacity analysis and energy saving of complex chemical processes using LSTM based on attention mechanism

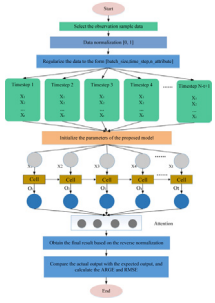
Yongming Han[a,b], Chenyu Fan[a,b], Meng Xu[a,b], Zhiqiang Geng[a,b,*], Yanhua Zhong[c]

[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
[c] Jiangmen Polytechnic, Jiangmen, Guangdong 529020, China

HIGHLIGHTS

- The AM-LSTM method is proposed.
- The AM-LSTM method is effectively verified through the UCI dataset.
- Production capacity analysis and energy saving model of complex chemical process is obtained.
- This proposed model is efficient in production prediction and energy saving of complex chemical processes.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Deep learning
LSTM
Attention mechanism
Energy saving
Production capacity analysis
Complex chemical processes

ABSTRACT

The production data of complex chemical processes are multi-dimensional, uncertain and noisy, and it is difficult to directly control raw materials consumption and measure the product quality. Therefore, this paper proposes a production capacity analysis and energy saving model using long short-term memory (LSTM) based on attention mechanism (AM) (AM-LSTM). The weights of the results sequence in the hidden layer, which have great influence on final results in the output layer, are calculated by the AM. Then the production prediction model is built using the LSTM to extract features of the input data and multiple time series results of the hidden layer. Compared with the common LSTM, the multi-layer perceptron (MLP) and the extreme learning machine (ELM), the applicability and the effectiveness of the proposed model is validated based on University of California Irvine repository (UCI) datasets. Finally, the proposed model is applied to analyze the production capacity and the energy saving potential of the purified terephthalic acid (PTA) solvent system and the ethylene production system of the complex chemical process. The experimental results verify the practicability and accuracy of the proposed model. Furthermore, the results offer the operation guidance for production capacity improvement through saving energy and reducing the energy consumption.

1. Introduction

The chemical industry occupies a very important position in social and economic development in China. The Purified terephthalic acid (PTA) process and the ethylene production process make important constitutes of the chemical industry. The PTA process, which is widely used in various fields like the light industry, chemical fiber, electronics, medicine, food and construction, serves as an important chemical raw

* Corresponding author at: College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, China.
E-mail address: gengzhiqiang@mail.buct.edu.cn (Z. Geng).

https://doi.org/10.1016/j.applthermaleng.2019.114072
Received 14 July 2018; Received in revised form 18 June 2019; Accepted 2 July 2019
Available online 04 July 2019
1359-4311/

material for production of the polytrimethylene terephthalate (PTT) and the polyethylene terephthalate (PET) [1]. The production capacity of the PTA plant in China is 46,890,000 tons per year, which occupies the most, and the domestic PTA capacity takes up 65.6% [2]. Because of the high cost of adding new production equipment, the overall energy efficiency is low [3]. Thus, how to reduce the acetic acid solvent consumption of the PTA plant has become a major research focus. The ethylene production industry, as a leader of the complex chemical process, has become one of the important signs to judge the level of the industrial development of a country [4,5]. According to the statistics data, the production of 1000,000 tons of ethylene requires about 3200,000 tons of petroleum, of which 18% (about 576,000 tons) is consumed by processing energy [6]. The ethylene plant has a direct impact on realizing goals of energy saving, low carbon emission and green gross domestic product (GDP). Therefore, a LSTM model is proposed in this paper, based on the attention mechanism (AM-LSTM), to analyse the production capacity and energy saving potentian of complex chemical processes.

The weights of the result sequence in the hidden layer, calculated by the AM, have great influence on final results in the output layer. Then the production prediction model is built using the LSTM to extract features of the input data and multiple time series results of the hidden layer. The applicability and the effectiveness of the proposed model is validated based on the university of California Irvine repository (UCI) dataset. Meanwhile, the validation results are compared with those of the common LSTM, the multi-layer perceptron (MLP) and the extreme learning machine (ELM). Finally, the production capacity analysis and energy saving models of the PTA solvent system and the ethylene production system in complex chemical processes are built. The experimental results verify the practicability and accuracy of the proposed model. Furthermore, the results offer the operation guidance for production capacity improvement through saving energy and reducing the energy consumption.

The organization of this paper is as follows: Section 2 presents the research status of energy efficiency analysis and energy saving of complex chemical processes. Section 3 describes the details of the AM-LSTM mode. And comparison with the BP, the ELM and the common LSTM based on the UCI standard dataset is presented in Section 3. Section 4 presents two practical study cases about the production capacity and energy saving analysis of the PTA solvent system and the ethylene production system. The conclusion is presented in Section 5.

2. Literature review

In recent years, many methods have been used to improve the energy efficiency of complex chemical processes. Due to the lack of energy consumption evaluation methods for existing ethylene industrial plants, Geng et al. [7] proposed a data variance fusion method for multivariate sequences based on the clustering analysis, and obtained the energy consumption index of the ethylene industrial plant. Han et al. [8] proposed a linear optimization fusion model based on fuzzy C-means (FCM) clustering to visually reveal the operation state and improve the energy efficiency of ethylene production plants. Gong et al. [9] proposed a new energy efficiency evaluation method for the ethylene production plant based on the DEA method, integrated with the factor analysis, under multiple conditions. Han et al. [10] proposed a performance analysis and energy optimization method of complex chemical processes using the data envelopment analysis cross-model (DEACM), integrated with the affinity propagation. Geng et al. [11] proposed an analysis method based on the improved environmental DEACM method to analyse and evaluate the energy and environmental efficiency of complex chemical processes. Dong et al. [12] located the fault in the chemical process, based on the principal components analysis (PCA) method. He et al. [13] proposed an improved soft measurement model based on the ELM, integrated with the PCA, to achieve on-line accurate measurement production. Zhu et al. [14] proposed a PCA and DEA integrated method for analysing the energy efficiency of ethylene production plants. In addition, Tian et al. [15] utilized the analytic hierarchy process (AHP) to evaluate the comprehensive environmental risk of the chemical industrial plant. Geng et al. [16] proposed the fuzzy C-Means-integrating- AHP (FAHP) to analyse the energy status of the PTA solvent system and the ethylene production system. However, the traditional machine learning methods mentioned above are not easy to migrate based on certain features. Therefore, the artificial neural network (ANN) was proposed such as the extreme learning machine (ELM) [17,18]. Peng et al. [19] proposed a novel sample clustering extreme learning machine (SC-ELM) model to measure the quality index of a high-density polyethylene (HDPE) process in a chemical plant. And the self-organizing ELM method was proposed to improve the energy efficiency of complicated chemical processes [20]. Mohammadi [21] proposed a new ELM model to estimate the wind power density effectively. He et al. [22] proposed an auto-associative hierarchical neural network (AHNN) to explore the problems of monitoring chemical processes with large numbers of input parameters and focus on dealing with datasets in high dimension. Meanwhile, the ANN is used in the heating energy demand and consumption reduction [23], energy efficiency indices prediction of driven wheels [24], control of the intelligent home system [25], and optimization of energy efficiency [26]. In recent years, the deep learning method has also been widely concerned and studied. Many models based on the convolutional neural network (CNN) [27] have achieved good results. Wu et al. [28] proposed a DCNN model consisting of convolutional layers, pooling layers, dropout, and fully connected layers for fault diagnosis in chemical processes. Lee et al. [29] proposed a new CNN model for fault detection and classification (FDC) in the field of semiconductor manufacturing processes. However, all the above methods cannot make use of the time series characteristics of chemical data. The recurrent neural network (RNN) [30] and its improved version with the long short-term memory (LSTM) [31] were designed specifically for processing the sequential data and could make full use of data features to achieve notable results in prediction. Ke et al. [32] introduced a deep neural network structure based on the LSTM as a soft sensor method to deal with strong nonlinearity and dynamics of chemical industrial processes. Zhao et al. [33] proposed a fault diagnosis method based on the LSTM to better separate different chemical process faults and provide more promising fault diagnosis performance. In addition, the LSTM was used to generate molecules in the drug discovery process [34] and energy load prediction [35]. Moreover, the chemical process output is often affected by one or several factors, some of which are even decisive. The performance can be improved by appropriately adjusting the weights to the inputs. Thus, the AM was proposed and mainly used in natural language understanding [36] and image recognition [37], but not used in the industrial process yet. Therefore, this paper proposes a novel AM-LSTM model to predict accurately the key variables of the complex chemical process data to reduce the energy consumption.

3. LSTM model based on attention mechanism

3.1. LSTM

The RNN is a loop neural network that the current output of a sequence is related to the previous output. Meanwhile, the RNN can remember the front information to the current calculation. And the input of the hidden layer includes the output of the input layer and the output of the hidden layer at the last timestep. A typical RNN is shown in Fig. 1.

It can be seen from Fig. 1 that the hidden layer and the next time result of the RNN are affected by the current and the previous result step by step. And the process of forward propagation can be obtained by Eqs. (1)–(3).

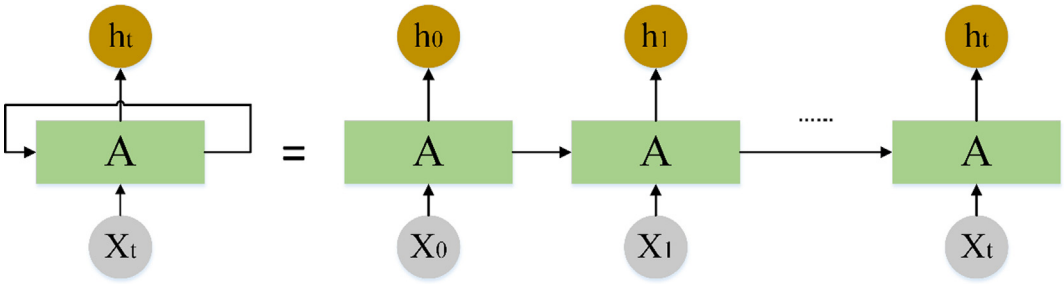


Fig. 1. The network structure of the RNN.

$$e_k^t = \sum_{h=1}^{H} w_{hk} b_h^t \tag{1}$$

$$a_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} b_{h'}^{t-1} \tag{2}$$

$$b_h^t = f(a_h^t) \tag{3}$$

where x_i^t is the i^{th} input data, a is the value of the weighted calculations, b represents the value calculated by the activation function, w_{hk}, w_{ih} and $w_{h'h}$ represent the weight matrix between the current hidden layer and the output, between the input and the current hidden layer and between the hidden layer status at the previous time and the current hidden layer, respectively. However, the traditional RNN encounters the gradient vanishing and the explosion. To solve the problem, the LSTM introduces a core element called cell. The structure of the cell in the LSTM is shown in Fig. 2.

The input-gate of the LSTM at t time is calculated by Eqs. (4) and (5).

$$a_l^t = \sum_{i=1}^{I} w_{il} x_i^t + \sum_{h=1}^{H} w_{hl} b_h^{t-1} + \sum_{c=1}^{C} w_{cl} s_c^{t-1} \tag{4}$$

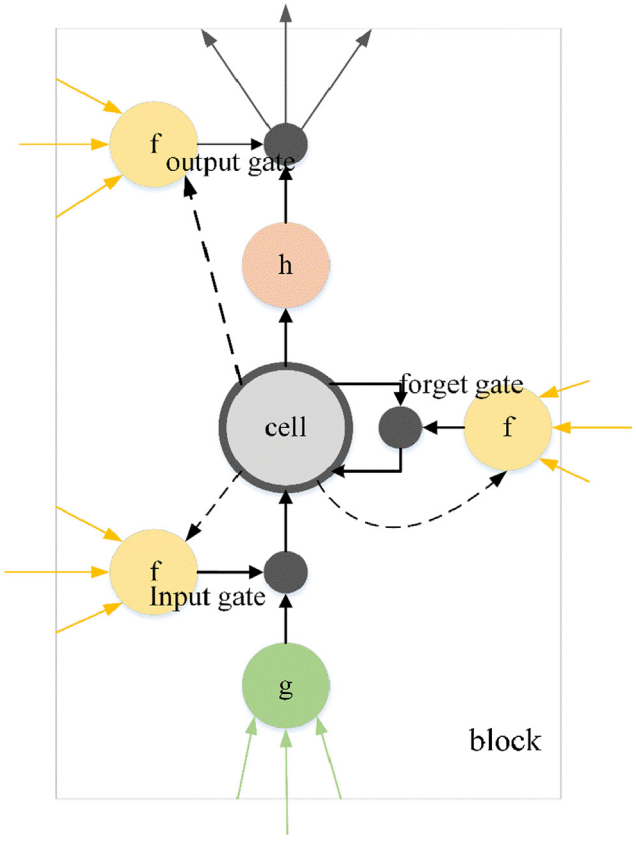


Fig. 2. The internal structure of the cell.

$$b_l^t = f(a_l^t) \tag{5}$$

The forget-gate can be obtained by Eqs. (6) and (7).

$$a_\varnothing^t = \sum_{i=1}^{I} w_{i\varnothing} x_i^t + \sum_{h=1}^{H} w_{h\varnothing} b_h^{t-1} + \sum_{c=1}^{C} w_{c\varnothing} s_c^{t-1} \tag{6}$$

$$b_\varnothing^t = f(a_\varnothing^t) \tag{7}$$

Due to the updating the cell state based on Eqs. (8) and (9), the output can be obtained by Eqs. (10) and (12).

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} b_h^{t-1} \tag{8}$$

$$s_c^t = b_\varnothing^t s_c^{t-1} + b_l^t g(a_c^t) \tag{9}$$

$$a_w^t = \sum_{i=1}^{I} w_{iw} x_i^t + \sum_{h=1}^{H} w_{hw} b_h^{t-1} + \sum_{c=1}^{C} w_{cw} s_c^t \tag{10}$$

$$b_w^t = f(a_w^t) \tag{11}$$

$$b_c^t = b_w^t h(s_c^t) \tag{12}$$

where in Eqs. (4)–(12), $x_i^t$is the i^{th} input data, s_c^t is the cell state at the current moment. s_c^{t-1} is the cell state at the last moment. All the weight matrix with w_{h*}orw_{*h} represent a general reference indicating any one side from the previous point to the current time. And w_{il} and $w_{i\varnothing}$ represents the weight matrix between input data and input-gate and between the input data and the forget-gate, respectively. w_{cl} and $w_{c\varnothing}$ represents the weight matrix between the cell state at the previous time and the input-gate and between the cell state at the previous time and the forget-gate, respectively. w_{iw} and w_{cw} represents the weight matrix between the input data and the output gate and between the Cell state at the previous time and the output gate, respectively. And the w_{ic} represents the original weight matrix of the input layer.

3.2. Attention mechanism

The attention mechanism can add the weight to the information of the previous part of the time or some concerned factors for different inputs and realizes the indirect association between the output and the related input. The attention mechanism is shown in Fig. 3.

And the attention mechanism can be described based on Eqs. (13)–(15).

$$u_t = tanh(w_w h_t + b_w) \tag{13}$$

$$\alpha_t = \frac{exp(u_t^T u_w)}{\sum_{k=1}^{T} exp(u_t^T u_w)} \tag{14}$$

$$s = \sum_t \alpha_t h_t \tag{15}$$

where w_w and u_w are the weight matrix of the attention mechanism, which indicate the information needed to be emphasized. u_t is the result of the first weighted calculation, b_w represents the bias of the attention mechanism, h_t represents the input of the attention mechanism, namely the output result of the hidden layer of the LSTM and α_t is the ultimate weight on h_t to get the final output s.

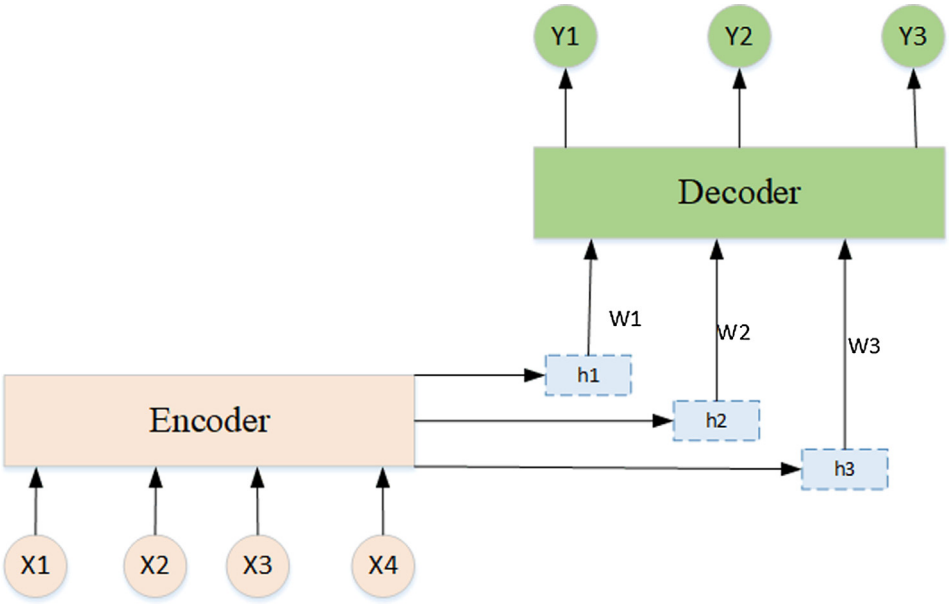


Fig. 3. The Attention Mechanism.

3.3. Evaluation criterion

In this paper, the root mean square error (RMSE) and the average relative error (ARGE) are used to evaluate the quality of the proposed model on the train data and the test data, which can be obtained by Eqs. (16) and (17).

$$\begin{cases} \begin{cases} RGE_i = Abs\left(\frac{NetOut_i^{inver} - ExpectOut_i^{inver}}{ExpectOut_i^{inver}}\right) \\ \quad if \quad ExpectOut_i^{inver} \neq 0 \qquad (a) \\ RGE_i = Abs\left(\frac{NetOut_i^{inver} - ExpectOut_i^{inver}}{ExpectOut_i^{inver}}\right) else \end{cases} \\ ARGE = \frac{\sum_{i=1}^{n} RGE_i}{N} * 100 \qquad (b) \end{cases} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (NetOut_i^{norm} - ExpectOut_i^{norm})^2}{N}} \tag{17}$$

where $ExpectOut_i^{inver}$ and $NetOut_i^{inver}$ $(i = 1, 2, 3 \cdots N)$ is the real output and the inverse normalized prediction value of all samples, and N is the total number of the samples.

3.4. The framework of the LSTM model based on the AM

The AM preserves the output of the LSTM at different time and selectively learns these inputs to associate them with the output sequence. Therefore, the probability of each item in the output sequence depends on the items selected in the input sequence. Given a training sample set $\{(X_n, Y_n) | n = 1, 2, 3 \cdots N; X_n \in R^{d1}; Y_n \in R^{d2}\}$, where X_n is the input data vector, Y_n is the real value of each sample, N is the total number of the samples. The algorithm steps of the LSTM model based on the AM are shown as the following:

Step 1: The input data X_n, Y_n are normalized to the scale of [0, 1] based on Eqs. (18)–(20).

$$\bar{x}_j^{(t)} = \begin{cases} \frac{x_j^{(t)} - X_j^{min}}{X_j^{max} - X_j^{min}}, \ if x_j^{max} \neq x_j^{min} \\ -1, \ other \end{cases} \tag{18}$$

Where $X_j^{max} = max\{x_j^{(1)}, x_j^{(2)} \cdots\cdots x_j^{(m)}\}$ (19)

$$X_j^{min} = min\{x_j^{(1)}, x_j^{(2)} \cdots\cdots x_j^{(m)}\} \tag{20}$$

Step 2: The data are regularized as the form of $[B, S, D]$, where B is the batch size, S is the sequence length and D is the number of the attribute. And then, the data of a dimension is shown as follows: $\begin{bmatrix} x_1^{(1)} & \cdots & x_1^{(D)} \\ \vdots & \ddots & \vdots \\ x_S^{(m)} & \cdots & x_S^{(D)} \end{bmatrix}$. From the first data X_0, the next S length of data are taken as the input data for the first time period, then from second data X_1, the next sequence length of data are taken as the input data for the second time period. Therefore, the N data can produce the input data of $N - step + 1$ time series altogether.

Step 3: The parameters of the proposed model are initialized.

Step 4: According to the Eqs. (4)–(12), the intermediate variables are calculated and updated to build the proposed model.

Step 5: The final output result of the proposed model can be calculated based on Eq. (21).

$$x_j^{(t)} = \bar{x}_j^{(t)} * (X_j^{max} - X_j^{min}) + X_j^{min} \tag{21}$$

where $\bar{x}_j^{(t)}$ represents the output of the proposed model and $x_j^{(t)}$ represents the final prediction value.

Step 6: Compared the actual output with the expected output, the ARGE and RMSE error of the proposed model are calculated.

The flow char of the proposed model is shown in Fig. 4.

3.5. UCI benchmark data set test

In order to verify the validity and the practicability of the proposed model, the Airfoil and the Wine Quality data sets from UCI standard datasets (For the details of the datasets, one can refer to the following website: http://archive.ics.uci.edu/ml/datasets.html) are selected to validate the ability for processing the regression problems, which are shown in Table 1. The standard data are normalized based on Eqs. (20)–(22) as the input of the BP, the ELM, the LSTM and the proposed model. The learning rate is set as 0.001, the momentum rate as 0.9 and training times as 3000. And then the RMSE and the ARGE are calculated to evaluate the prediction performance.

The results of these models based on two kinds of UCI standard data

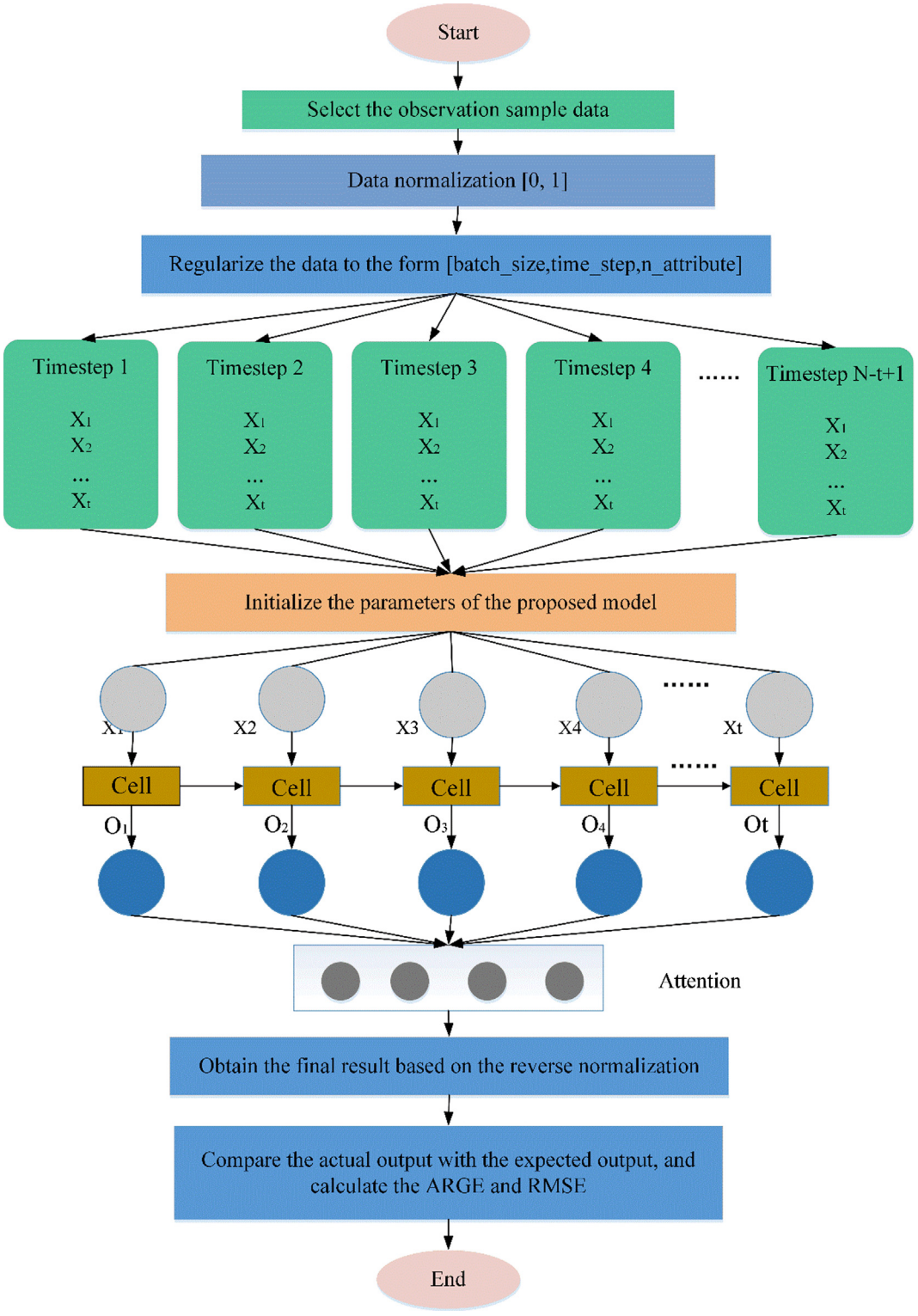


Fig. 4. The flow chart of the AM-LSTM.

Table 1
The specification of UCI data.

Data Set	Input Size	Output Size	Train Samples	Test Samples
Airfoil	6	1	1000	503
WineQuality	11	1	1000	599

Table 2
The testing result of Airfoil Data.

Model	RMSE	ARGE(%)
The BP	5.66	3.48
The ELM	22.1	4.97
The LSTM	4.42	2.64
The AM-LSTM	4.00	2.30

Table 3
The testing result of Wine-Quality Data.

Model	RMSE	ARGE(%)
The BP	0.83	12.06
The ELM	0.82	12.53
The LSTM	0.75	10.98
The AM-LSTM	0.71	10.57

sets are shown in Tables 2 and 3.

From Tables 2 and 3, the performance of the LSTM model is superior to the ELM and the BP. Compared with the common LSTM model, the AM-LSTM model has higher generalization accuracy, which indicates that the AM plays a role in enhancing the performance of the LSTM model under the same structure and the same test environment. And the whole results also show the validity and the practicability of the AM-LSTM model.

4. Case study: Productivity capacity analysis and energy saving in complex chemical processes

In order to improve the energy efficiency and save the energy of complex chemical processes, the acetic acid production in the PTA solvent system and the ethylene production in the ethylene production process are analysed and predicted by the proposed model. Complex chemical data has dynamic, real-time, continuous characteristics. And the traditional machine learning and neural network methods can extract main features, but these methods cannot utilize the temporal characteristics of complex chemical data. The LSTM network is better at processing the sequence data, so it can extract maim features and the timing characteristics of the chemical data by using the timing relationship between data at different times. Meanwhile, the production status at different times in the complex chemical production process may have different degrees of influence on the current production situation. Therefore, the attention mechanism assigns different weights of each output in the LSTM at different times to simulate the dependence among the chemical data. Based on the analysis above, this proposed method has certain feasibility and superiority in the productivity capacity analysis and energy saving of complex chemical data.

Table 4
The 17 input variables of the PTA production plant.

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

4.1. Production prediction and energy saving of PTA solvent systems

The PTA solvent system is a key part of the PTA production which consists of three parts: the solvent dehydration tower, the re-boiler and the reflux tank. The PTA production takes the *para*-xylene (PX) as the raw material and the Co-Mn-Br as the catalyst, the whole production process is mainly composed of slurry preparation, hydrogenation, crystallization, separation, drying and powder transportation. Finally, the fiber grade PTA is obtained. During the PTA production process, the consumption of the acetic acid is an important indicator to measure the efficiency of the PTA solvent system [38] and is also an important sign to measure the advanced technology used in the system equipment. Therefore, it is necessary to reduce the consumption of the acetic acid under the condition of ensuring the normal capacity of the PTA plant. The PTA solvent system is shown in Fig. 5.

The data from 253 time points collected by DCS of the PTA plant are selected. Based on analysis of the operating characteristic of the PTA production plant, there are 17 factors in the solvent system that affect the consumption of the acetic acid at the top of the tower [39]. These 17 factors correspond to the NO.1-NO.17 mentioned as shown in

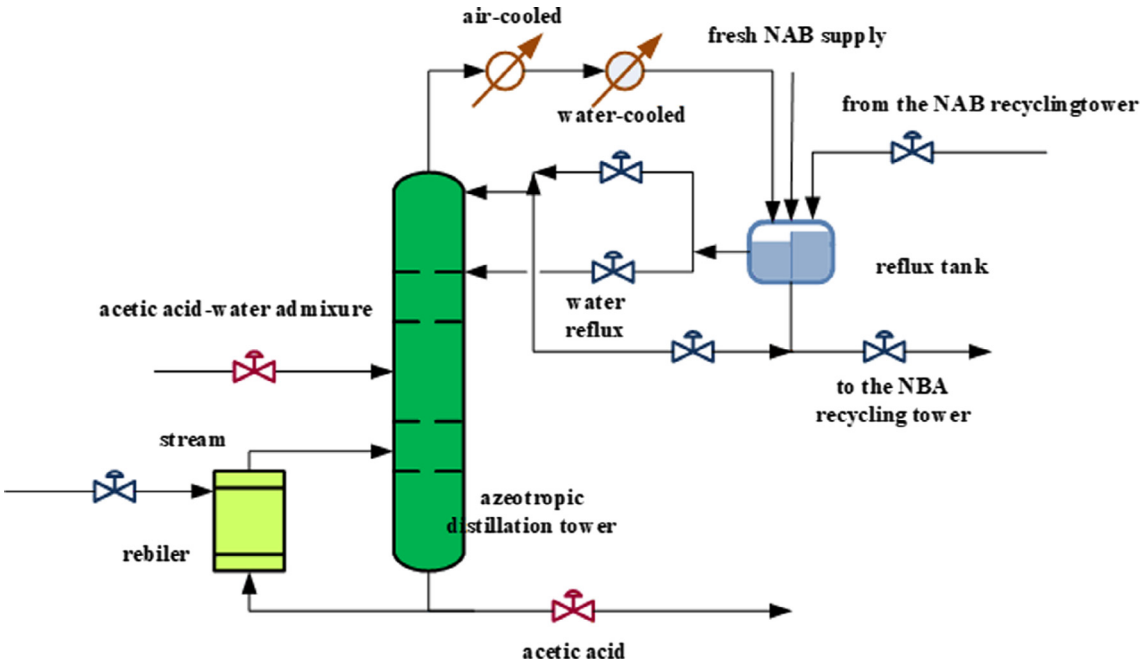


Fig. 5. The PTA solvent system.

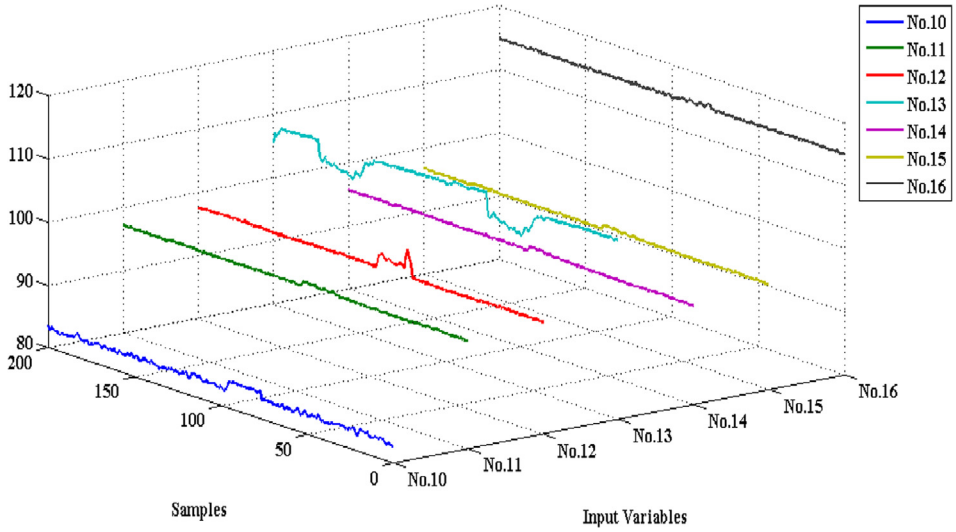


Fig. 6. The temperature variables of the PTA plant.

Table 5
Specification of PTA data sets.

#Samples		#Attributes	
Training	Testing	Inputs	Output
173	80	17	1

Table 4. And the temperature variables of the PTA plant in the actual production is shown in Fig. 6.

The 17 factors mentioned above are used as the input variables. Because the top of the acetic acid is difficult to measure, the electrical conductivity reflecting the change of the acetic acid content indirectly at the top of the tower are used as the output variable [13]. And 173 sets of training data, and 80 sets of test data are used as shown in Table 5. Then the input data are normalized to the scale of [0, 1], selecting the time _step (T) = 10, and batch_size = 8. Finally, the input data at every time period are regulated as the format [8,10,17] and [8,10,1], and the test data are processed the same way. The regulated data is considered as the input data of the LSTM to extract the feature and produce T time series output results which are used as the input of the AM. Then the AM selectively learns and associates these intermediate output results with the final output sequence to predict and analyze the PTA plant production and provide production guidance.

Table 6
The result of different model.

Model	Train Data		Test Data	
	RMSE	ARGE(%)	RMSE	ARGE(%)
The BP	0.469	0.814	0.476	0.836
The ELM	0.459	0.799	0.468	0.817
The LSTM	0.039	0.066	0.202	0.323
The AM-LSTM	0.027	0.042	0.107	0.158

Based on the PTA data, the relation between the number of hidden nodes and testing errors for the BP and the ELM are described in Fig. 7. It can be seen from Fig. 7 that the number of the hidden layer nodes of the BP and the ELM is supposed to be 35 and 40, respectively.

The ARGE and the RMSE of the training set and the testing set of the BP, the ELM, the LSTM and the AM-LSTM model are shown in Table 6.

From Table 6, the training and testing ARGE of the AM-LSTM is 0.042 and 0.158, respectively, which are better than the ELM. Meanwhile, the training and the testing RMSE of the AM-LSTM is 0.027 and 0.107, respectively, which are better than the ELM. The prediction results of the LSTM and the AM-LSTM are obviously better than the BP and the ELM. Compared with the testing RMSE and ARGE of the common LSTM is 0. 202 and 0.323, the result of the AM-LSTM is better under the same test environment, which means that the AM plays a role

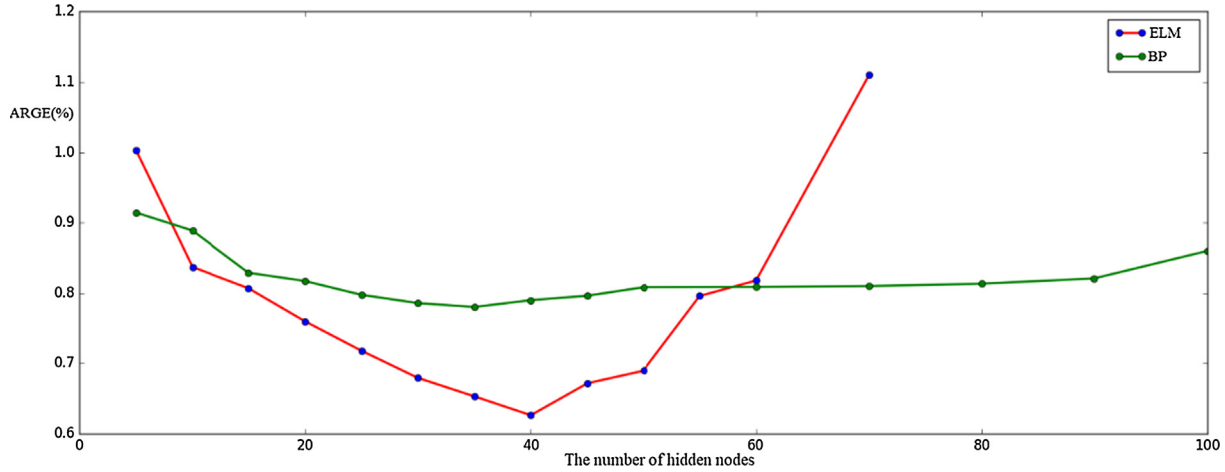


Fig. 7. Relation chart between error of testing and number of hidden neurons based on the PTA data.

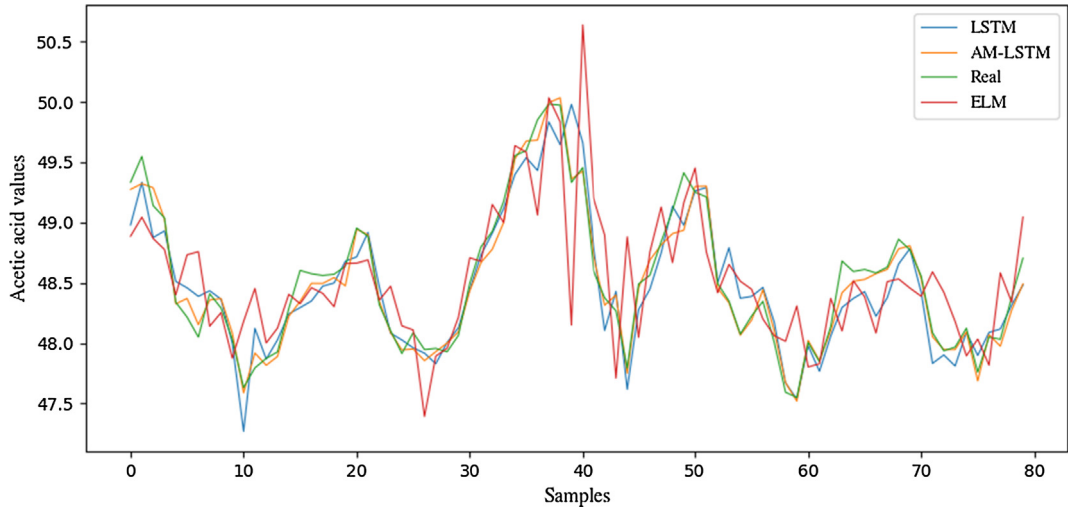


Fig. 8. The prediction values obtained by different models and the real values of the acetic acid.

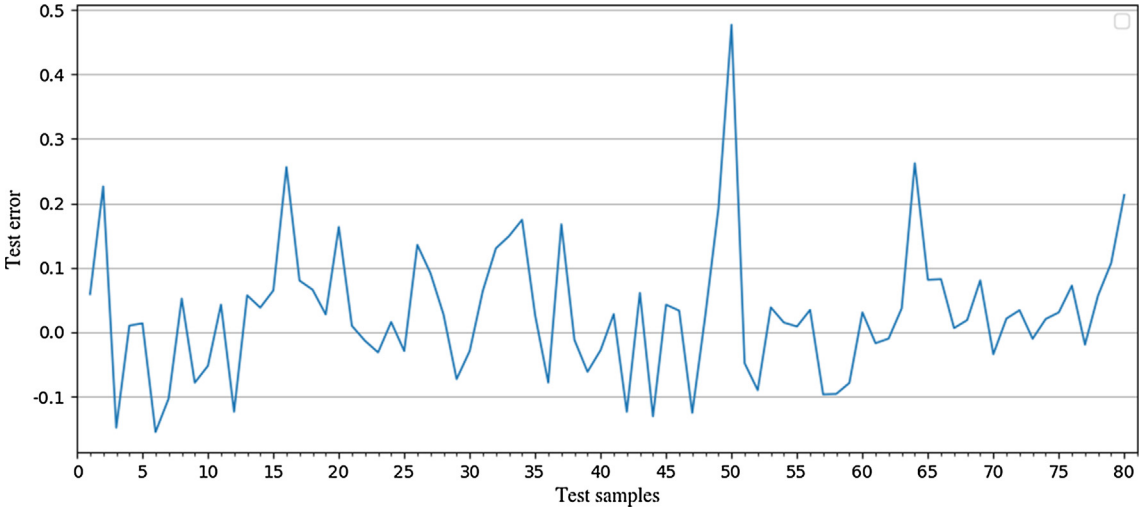


Fig. 9. The difference between the true value and the prediction result of the PTA plant.

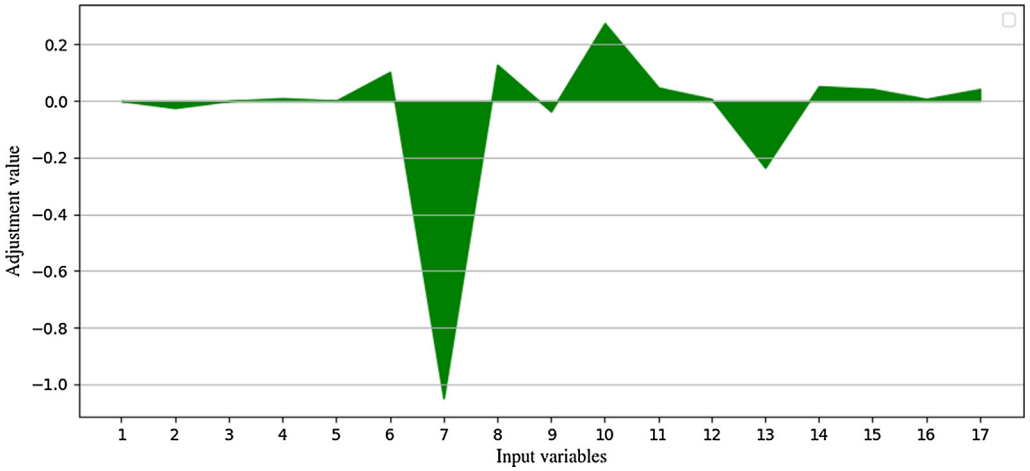


Fig. 10. The adjustment of production parameters of 50th sample.

in enhancing the accuracy of the LSTM in the PTA production process. The prediction values obtained by common LSTM (common_lstm), AM-LSTM (attention_lstm), ELM (elm_data) and the real values (real_data) of the acetic acid are shown in Fig. 8.

It can be seen from the Fig. 8 that the data fitting of the proposed model is better and more stable, and a fairly accurate prediction value of the acetic acid consumption in the PTA plant is obtained, which can be used to estimate and analyze the energy efficiency level and guide

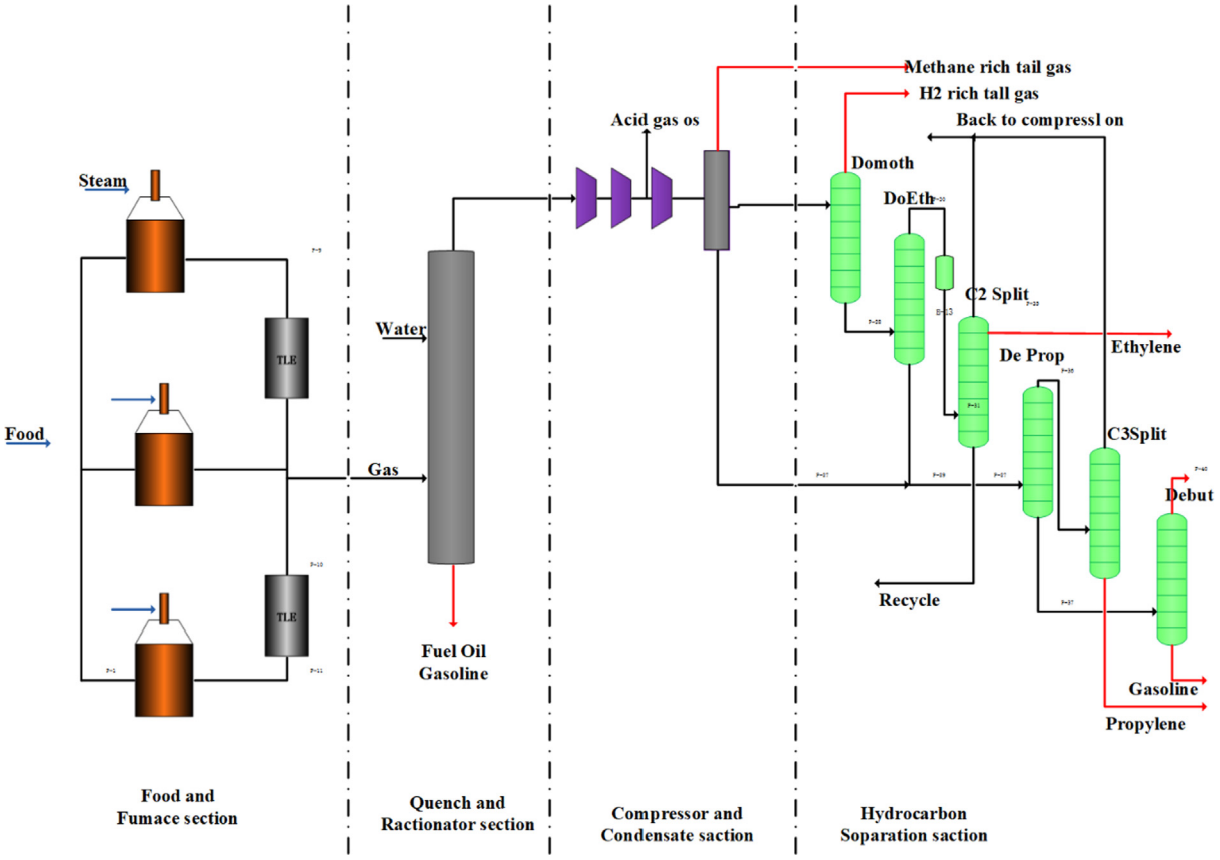


Fig. 11. The ethylene production plant.

Table 7
The specification of the data.

	Train Data	Test Data
Samples	228	100
Num of input	5	5
Num of output	1	1
Input attributes	consumption of crude oils, fuel, water, steams and electricity	
Output attribute	ethylene production	

the PTA production. The gap between the true value and the predicted value obtained by the AM-LSTM is shown in Fig. 9, in which the x-axis represents the total of the 80 samples of the predicted samples, and the y-axis represents the difference between the real and the predictive values $Y_{real} - Y_{pred}$.

According to the data of group 2, 16, 50 and 64, the actual consumption of the acetic acid is obviously larger than the prediction output of the proposed model. For example, the actual electrical conductivity of 50th test data is 49.4216 and the prediction value of the

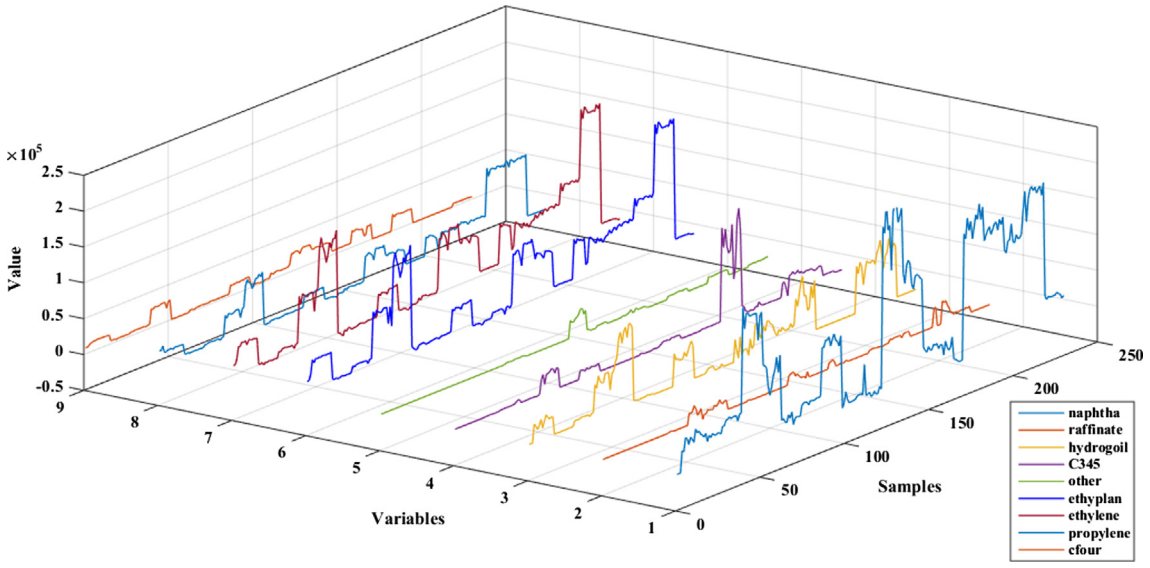


Fig. 12. The important raw materials.

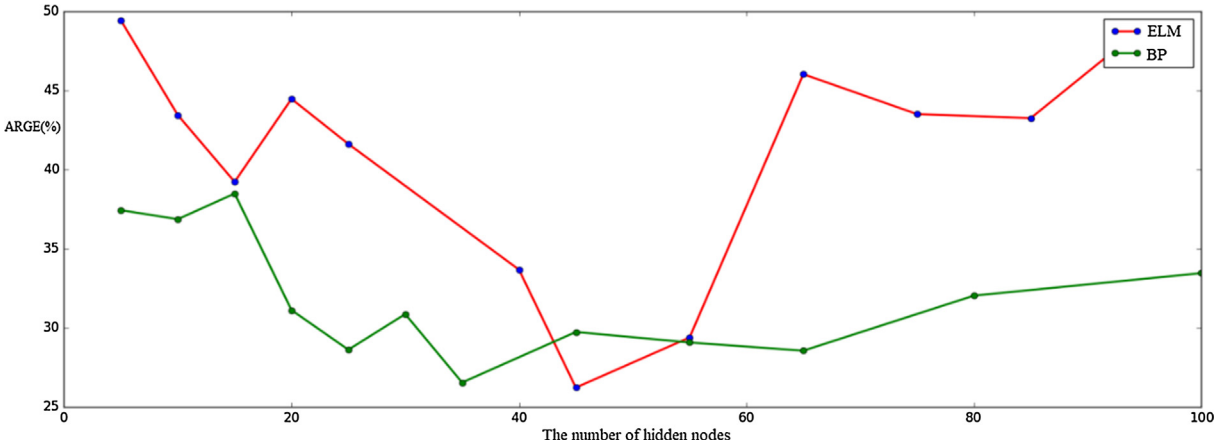


Fig. 13. Relation chart between error of testing and number of hidden neurons based on the ethylene data.

Table 8
The results of different models.

	Train Data	Test Data
Model	ARGE (%)	ARGE (%)
The BP	24.34	26.80
The ELM	15.85	26.25
The LSTM	5.93	11.24
The AM-LSTM	2.81	4.08

proposed model is 48.9358, which shows that the PTA plant at this moment operates without full load in the PTA manufacturing process and the energy utilization rate is low. However, according to the data of group 4, 5, 11, 21, 22, 23, 53, 54, and 61, the real value is very close to and even lower than the prediction value, indicating that the operation status of the PTA plant at this moment is more perfect and the energy efficiency level is high. Therefore, the future production scale can be adjusted to improve the energy efficiency level and reduce the consumption with reference to the above eight time points in the complex chemical production processes. For low energy utilization samples, when $Y_{real} > Y_{pred}$ exists, and the true value Y'_{real} is close to the predicted value Y'_{pred} and Y_{pred}, the device has a high energy efficiency level by adjusting the Y_{real}input based on the 17 input variables of Y'_{real}. For example, in order to achieve the goal of energy saving, the changes of the production parameters for the 50th test data are shown in Fig. 10.

It can be seen from Fig. 10, in order to product more effectively and increase the energy utilization, referring to the 21st group of data, the parameters NO.1, NO.4, NO.5, NO.6, NO.8, NO.10-NO.12, NO.14-NO.17 should be increased 0.00026, 0.009, 0.00104, 0.10217, 0.12757, 0.27451, 0.04719, 0.00686, 0.05157, 0.04198, 0.00685, 0.04174 respectively. Meanwhile, the parameters NO.2, NO.7, NO.9 and NO.13 should be reduced 0.024486, 1.04989, 0.03689 and 0.2354, respectively. According to the above analysis and adjustment, the PTA plant can reduce the consumption of the acetic acid about 0.459.

4.2. Production prediction and energy saving analysis of the ethylene industrial process

The ethylene production process can be divided into two parts: cracking and separation. The main process of cracking is to process raw materials such as natural gas, refinery gas, crude oil and naphtha into cracking gas and provide them to other ethylene plants, and the process of separation consists of three steps: rapid cooling, compression and final separation part [40]. The ethylene industry standard DB 37/751-2007 and GB/T 2589-2008 [41,42] are adopted to perform the division of ethylene production plants battery limitation. The ethylene production plant is shown in Fig. 11. It is known from the energy battery limit of the ethylene production including: the crude oils including naphtha, light diesel fuel, raffinate, hydrogenation tail oil, light hydrogenation tail oil, C345 and others, fuel including light weight oil, heavy weight oil and fuel gas, steams including super high-pressure steam, high-pressure steam, medium pressure steam and low-pressure steam, water including circulating water, industrial water, boiler water and other water, and the electricity [40,43].

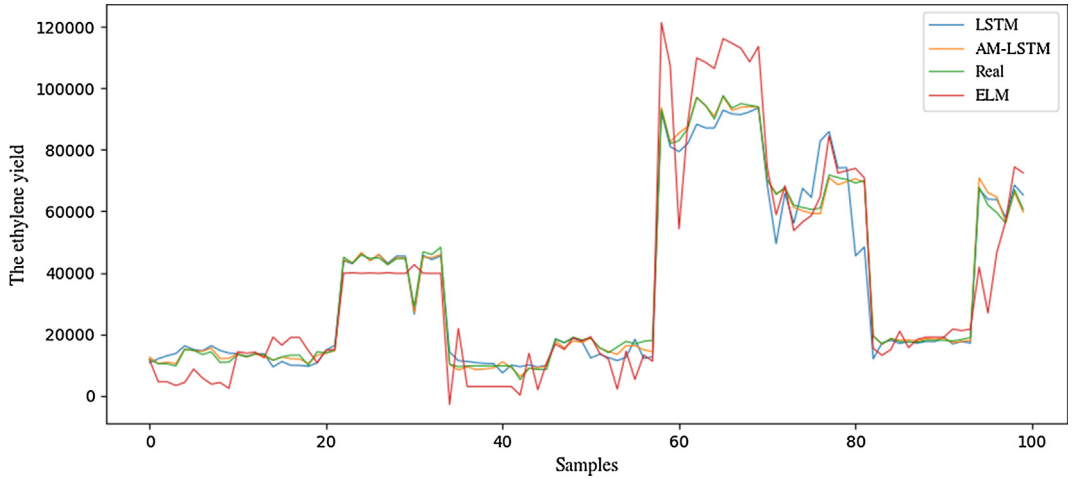


Fig. 14. The prediction values obtained by different models and the real values of the ethylene production.

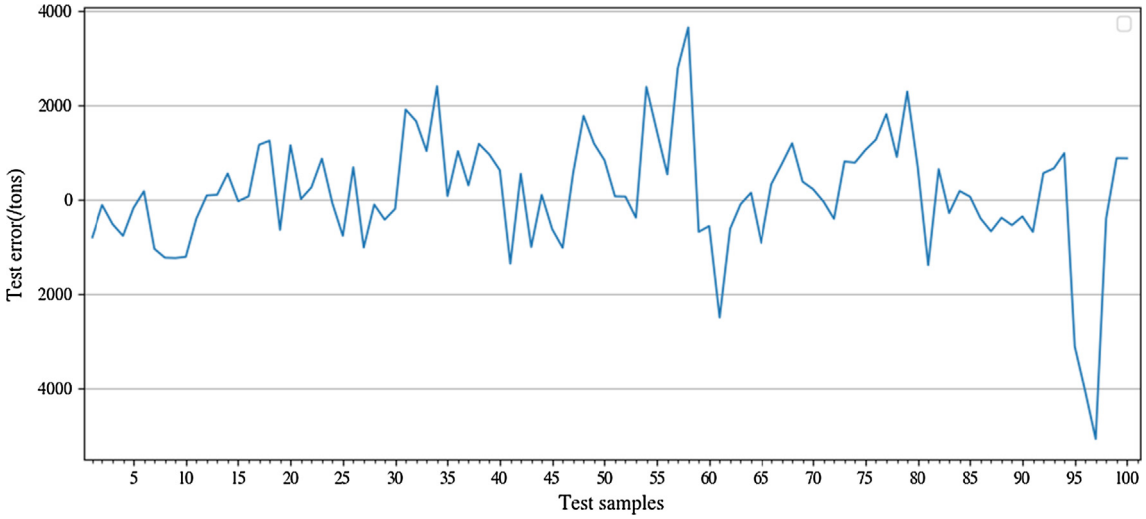


Fig. 15. The difference between the true value and the prediction result of the ethylene plant.

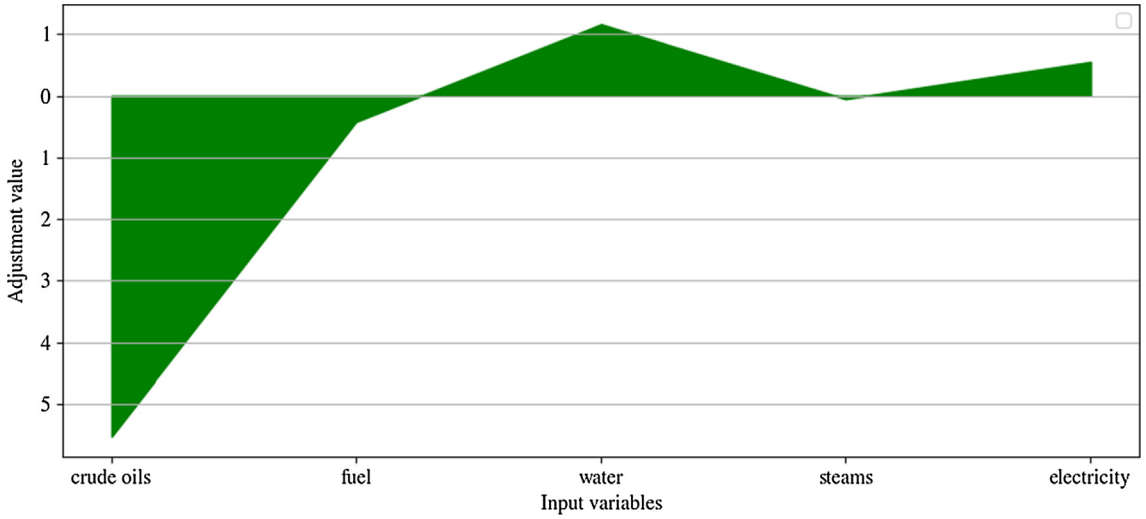


Fig. 16. The adjustment of production parameters of 97th test data.

The actual production data of an ethylene plant with the same technology from 2011 to 2012 is used in this paper [40]. The consumption of crude oils, fuel, water, steams and the electricity are used as the inputs of the proposed model, and the yield of the ethylene production is used as the output indexes [39]. The units of the raw material are Ton, and the units of other inputs are all converted to GJ/Ton(GJ) ethylene [43]. And 228 sets of actual statistical data in 2011 are used as the training data and 100 sets of ethylene production in 2012 are used as the test data. The specification of the data is shown as Table 7. And the consumption data of some important raw materials are shown in the Fig. 12.

The proposed model obtains the production prediction and energy saving model of the ethylene plant. The input data are normalized to the scale of [0, 1], and then selecting the time_step(T) = 10, and batch_size = 10. The input data of training and testing at every time period is regulated as the format [10,10,5] and [10,10,1]. The regulated data is considered as the input data of the LSTM to extract the feature and produce T time series output results which are used as the input of the AM. Then the AM selectively learns and associates these intermediate output results with the final output sequence to predict and analyze the ethylene production and provide the production guidance.

Based on the experimental data, the relation between the number of hidden nodes and testing errors for the BP and the ELM is shown in Fig. 13.

It can be seen from Fig. 13 that the number of hidden layer nodes of the BP and the ELM are supposed to be 35 and 45, respectively. For the BP, the ELM, the LSTM and the AM-LSTM, the ARGE of the training set and the testing set is calculated and shown in the Table 8. Because the RMSE result is related to the anti-normalization process, and the ethylene data is relatively large, so there is no comparison here. Moreover, the ARGE is more representative.

According to the Table 8, the ARGE error of the AM-LSTM on the test data is 4.08% which is lower than that of the BP and the ELM. And the prediction results of the LSTM and the AM-LSTM are obviously more advantageous than the BP and the ELM. Compared with the training and testing ARGE of the common LSTM is 5.93% and 11.24%, the result of the AM-LSTM has better performance under the same test environment, which means that the AM plays a role in enhancing the accuracy of the LSTM in the ethylene production process. The prediction values obtained by the common LSTM (common_lstm), the AM-LSTM (attention_lstm), the ELM (elm_data) and the real values (real_data) of the ethylene yield are shown in Fig. 14.

It can be seen from the Fig. 14 that the data fitting of the AM-LSTM model is better and relatively stable. Moreover, a fairly accurate prediction value of the ethylene production can be obtained to estimate

and discriminate the energy efficiency level and guide the ethylene production. The gap between the true value and the predicted value obtained by the AM-LSTM is shown in Fig. 15, in which the x-axis represents the total of the 100 groups of the test samples, and the y-axis represent the difference between the real and the predicted values $Y_{real} - Y_{pred}$.

According to the data of group 7, 8, 10, 27, 43, 61 and 97, the actual ethylene yield is obviously lower than the prediction output of the proposed model. And in the ethylene production process, the ethylene yield is low under the same original input conditions, indicating that the energy utilization rate is relatively low [40]. For example, the actual ethylene yield of 97th is 59,582 Tons and the prediction value of the proposed model is 64,655 Tons which indicates that the load of the ethylene production plant is low and the energy utilization rate is low at this time. However, according to the data of group 2, 5, 6, 12 to 16, 21 to 24, 35 and 37, the real values are close to and even higher than the prediction value, which indicates that the production efficiency at these series is high, so the production of other series can be adjusted to improve the production efficiency according to these production series in the ethylene production plant. For low energy utilization samples, when $Y_{real} < Y_{pred}$ exists, and the true value Y'_{real} is close to the predicted value Y'_{pred} and Y_{pred}, the device has a high energy efficiency level by adjusting the Y_{real}input based on the 5 input variables of Y'_{real}. For example, to further improve the production efficiency of 97th test data, the changes of the production parameters for the 97th test data are shown in the Fig. 16.

It can be seen from the Fig. 16, to improve the ethylene production and the energy utilization, refer to the optimal allocation of 72th test data, the parameters consumption of crude oils, fuel and steams should be increased 553 Tons, 0.42 GJ and 0.05 GJ, respectively. Meanwhile, the parameters water and electricity should be reduced 1.16 GJ and 0.55 GJ, respectively. According to the above analysis and adjustment, the ethylene yield can be increased by about 5884 Tons.

5. Conclusion

This paper proposes an improved AM-LSTM model. The AM can calculate the weights of the result sequence in the hidden layer. Then the production prediction model is built by the AM-LSTM. The applicability and the effectiveness of the proposed model is validated based on the university of California Irvine repository (UCI) dataset. Meanwhile, the validation results are compared with those of the BP, the ELM and the LSTM. Finally, the proposed model is used to build the production capacity analysis and energy saving model of the complex chemical process. The experimental results show that the proposed model can reduce the acetic acid production and improve the ethylene yield by reasonably configuring the raw materials in the PTA production process and the ethylene plant. However, compared with the traditional machine learning method, the proposed model has higher complexity and. As the time series of data grows, the gradient may disappear and the accuracy of the proposed model may been reduced. Meanwhile, the computation time and cost have also been increased.

In the future research, we will further study and discuss the application and comparative result of the multi-layer LSTM model based on the AM and the bi-directional LSTM model in the chemical process. Moreover, the carbon emission index is introduced to realize the real-time online analysis of the chemical data.

Acknowledgement

This work is partly financial supported by the National Key Research and Development Program of China (2018YFB0803501), National Natural Science Foundation of China (61603025 and 61673046) and the Fundamental Research Funds for the Central Universities (XK1802-4).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.applthermaleng.2019.114072.

References

[1] P. Huang, Current status of PTA industry in China, Synth. Technol. Appl. (2014).
[2] Y.H. Wang, Analysis of PTA market in China, Poly. Ind. (2016).
[3] Z.Q. Geng, J. Chen, Y.M. Han, Energy efficiency prediction based on PCA-FRBF model: a case study of ethylene industries, IEEE Transac. Syst. Man Cybernet. Syst. 47 (8) (2017) 1763–1773.
[4] G.F. Ma, Y.H. Xu, X. Guo, Review of sinopec's ethylene production in 2014, Ethyl. Ind. 1 (2015) 1–5.
[5] R.W. Yu, G.F. Ma, Y.H. Xu, Review of sinopec's ethylene production in 2015, Ethyl. Ind. (2016).
[6] C. Yu, X. Gu, Z. Geng, Ethylene industry efficiency analysis based on EPI, Ciesc J. 63 (9) (2012) 2931–2935.
[7] Geng Z Q, Han Y M, Zhang Y, Shi X Y. Data fusion-based extraction method of energy consumption index for the ethylene industry. Life System Modeling and Intelligent Computing, International Conference on Life System Modeling and Simulation, Lsms 2010, and International Conference on Intelligent Computing for Sustainable Energy and Environment, Icsee 2010, Wuxi, China, September 17-20, 2010, Proceedings. DBLP, 2010:pp. 84–92.
[8] Y.M. Han, Z.Q. Geng, Y.X. Qu, Q.X. Zhu, Linear optimization fusion model based on fuzzy C-means: case study of energy efficiency evaluation in ethylene product plants, J. Anal. Appl. Pyrol. 125 (2017) 347–355.
[9] S.X. Gong, S. Cheng, Z. Li, Energy efficiency evaluation based on DEA integrated factor analysis in ethylene production, Chin. J. Chem. Eng. 25 (6) (2017) 793–799.
[10] Y.M. Han, C. Long, Z.Q. Geng, Q.X. Zhu, Y.H. Zhong, A novel DEACM integrating affinity propagation for performance evaluation and energy optimization modeling: application to complex petrochemical industries, Energy Convers. Manage. 183 (2019) 349–359.
[11] Z.Q. Geng, J.G. Dong, Y.M. Han, Q.X. Zhu, Energy and environment efficiency analysis based on an improved environment DEA cross-model: case study of complex chemical processes, Appl. Energy 205 (2017) 465–476.
[12] G.S. Dong, W. Ren, A real-time fault monitoring system for the dynamic chemical process based on PCA, Comput. Appl. Chem. 27 (10) (2010) 1357–1360.
[13] Y.L. He, X. Wang, Q.X. Zhu, Modeling of acetic acid content in purified terephthalic acid solvent column using principal component analysis based improved extreme learning machine, Contr. Theory Appl. 32 (1) (2015) 80–85.
[14] Q.X. Zhu, X. Chen, Y.L. He, X.B. Gu, Energy efficiency analysis for ethylene plant based on PCA-DEA, CIESC J. 66 (1) (2015) 278–283.
[15] Y. Tian, Environmental risk assessment of chemical industry park based on analytic hierarchy process, International Symposium on Social Science, (2015).
[16] Z.Q. Geng, L. Qin, Y.M. Han, Q.X. Zhu, Energy saving and prediction modeling of petrochemical industries: a novel ELM based on FAHP, Energy 122 (2017) 350–362.
[17] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, IEEE International Joint Conference on Neural Networks, 2004. Proceedings. IEEE, 2004, pp. 985–990.
[18] J. Tang, C. Deng, G.B. Huang, Extreme learning machine for multilayer perceptron, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2017) 809–821.
[19] D. Peng, Y. Xu, Y. Wang, Soft-sensing in complex chemical process based on a sample clustering extreme learning machine model, IFAC PapersOnLine 48 (8) (2015) 801–806.
[20] Z.Q. Geng, J.G. Dong, J. Chen, Y.M. Han, A new self-organizing extreme learning machine soft sensor model and its applications in complicated chemical processes, Eng. Appl. Artif. Intell. 62 (C) (2017) 38–50.
[21] K. Mohammadi, S. Shamshirband, P.L. Yee, D. Petkovic, M. Zamani, Predicting the wind power density based upon extreme learning machine, Energy 86 (2015) 232–239.
[22] Y.L. He, Y. Xu, Z.Q. Geng, Q.X. Zhu, Soft sensor of chemical processes with large numbers of input parameters using auto-associative hierarchical neural network, CIESC J. 23 (1) (2015) 138–145.
[23] G. Jahedi, M.M. Ardehali, Wavelet based artificial neural network applied for energy efficiency enhancement of decoupled HVAC system, Energy Convers. Manage. 54 (1) (2012) 47–56.
[24] H. Taghavifar, A. Mardani, Applying a supervised ANN (artificial neural network) approach to the prognostication of driven wheel energy efficiency indices, Energy 68 (8) (2014) 651–657.
[25] T. Teich, F. Roessler, D. Kretz, S. Franke, Design of a prototype neural network for smart homes and energy efficiency, Procedia Eng. 69 (1) (2014) 603–608.
[26] I. Monedero, F. Biscarri, C. León, Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant, Exp. Syst. Appl. 39 (10) (2012) 9860–9867.
[27] J. Bouvrie, Notes on convolutional neural networks, Neural Nets (2006).
[28] H. Wu, J. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, Comput. Chem. Eng. (2018).
[29] K.B. Lee, S. Cheon, O.K. Chang, A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes, IEEE Trans. Semicond. Manuf. 30 (2) (2017) 135–142.
[30] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, Recurrent neural network based language model, Eleventh Annual Conference of the International Speech Communication Association, (2010).

[31] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[32] W.S. Ke, D.X. Huang, F. Yang, Y.H. Jiang, Soft sensor development and applications based on LSTM in deep neural networks, Comput. Intell. IEEE (2018) 1–6.
[33] H. Zhao, S. Sun, B. Jin, Sequential fault diagnosis based on LSTM neural network, IEEE Access 99 (2018) 1.
[34] Ertl P, Lewis R, Martin E. In silico generation of novel, drug-like chemical matter using the LSTM neural network. 2018.
[35] D.L. Marino, K. Amarasinghe, M. Manic, Building energy load forecasting using deep neural networks, Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE. IEEE, 2016, pp. 7046–7051.
[36] L. Wang, Z. Cao, G.D. Melo, Z. Liu, Relation classification via multi-level attention CNNs, Meeting of the Association for Computational Linguistics, 2016, pp. 1298–1307.
[37] J. Liu, G. Wang, P. Hu, L.Y. Duan, A.C. Kot, Global context-aware attention LSTM networks for 3D action recognition, IEEE Conf. Comp. Vision Pattern Recogn. IEEE Comp. Soc. (2017) 3671–3680.
[38] Y.L. He, Z.Q. Geng, Q.X. Zhu, Data driven soft sensor development for complex chemical processes using extreme learning machine, Chem. Eng. Res. Des. 102 (2015) 1–11.
[39] Y.M. Han, H. Wu, M.H. Jia, Z.Q. Geng, Y.H. Zhong, Production capacity analysis and energy optimization of complex petrochemical industries using novel extreme learning machine integrating affinity propagation, Energy Convers. Manage. 180 (2019) 240–249.
[40] Y.M. Han, Z.Q. Geng, Q.X. Zhu, Y.X. Qu, Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry, Energy 83 (2015) 685–695.
[41] China standards: the general computing guide of special energy consumption (GB/T2589-2008). 2008.
[42] China standards: the limitation of energy consumption for ethylene product (DB37/751-2007). 2008.
[43] Y.M. Han, Q.X. Zhu, Z.Q. Geng, Y. Xu, 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. 115 (2017) 280–291.