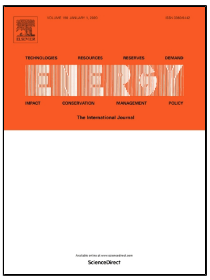

# Energy optimization and prediction modeling of petrochemical industries: An improved convolutional neural network based on cross-feature

Zhiqiang Geng, Yanhui Zhang, Chengfei Li, Yongming Han, Yunfei Cui, Bin Yu


| | |
|---|---|
| PII: | S0360-5442(19)32546-0 |
| DOI: | https://doi.org/10.1016/j.energy.2019.116851 |
| Reference: | EGY 116851 |
| To appear in: | *Energy* |
| Received Date: | 26 May 2019 |
| Accepted Date: | 24 December 2019 |

Please cite this article as: Zhiqiang Geng, Yanhui Zhang, Chengfei Li, Yongming Han, Yunfei Cui, Bin Yu, Energy optimization and prediction modeling of petrochemical industries: An improved convolutional neural network based on cross-feature, *Energy* (2019), https://doi.org/10.1016/j.energy.2019.116851

# Energy optimization and prediction modeling of petrochemical industries: An improved convolutional neural network based on cross-feature

Zhiqiang Geng[1,3], Yanhui Zhang[1,3], Chengfei Li[2], Yongming Han[1,3,*] , Yunfei Cui[1,3], Bin Yu[4]

[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;

[2] Department of Intelligent Manufacturing, Wu Yi University, Guangdong, Jiangmen 529020, China;

[3] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China

[4] Hengli Petrochemical co. LTD, Liaoning, Dalian 116000, China

**Abstract:** The petrochemical industry is the top priority of the national economy and sustainable development. For the purpose of improving the energy efficiency in the petrochemical industry, an energy optimization and prediction model based on the improved convolutional neural network (CNN) integrating the cross-feature (CF) (CF-CNN) is proposed. The CF can combine the correlation between features to obtain the input of the CNN, which can avoid over-fitting problems caused by fewer features. Then the CNN is designed as a three-layer structure and the Rectified Linear Unit (ReLU) is introduced to achieve better generalization capability and stability with boiler fluctuations in the petrochemical industry. The developed method has better performances of modeling accuracy and applicability than that of the back-propagation (BP) neural network and the extreme learning machine (ELM) on University of California Irvine (UCI) benchmark datasets. Furthermore, the developed method is applied to establish an energy optimization and prediction model of ethylene production systems in the petrochemical industry. The experimental results testify the capability of the proposed method. Meanwhile, the average relative generalization error is 2.86%, and the

* Corresponding author.
E-mail address: hanym@mail.buct.edu.cn (Yongming, Han).

energy utilization efficiency increases by 6.38%, which leads to reduction of the carbon emissions by 5.29%.

**Key words:** Production prediction modeling; Energy optimization; Carbon emissions reduction; Convolutional neural network; Cross-feature; Petrochemical industry

## 1. Introduction

The petrochemical industry is the top priority of the national economy and sustainable development. And the ethylene industry is a key part of complex petrochemical production industry. Currently, the ethylene is one of the demanding basic chemical materials [1] and the ethylene production consumes more than 15% of energy (including fuels and materials) in thousands of chemical products [2]. However, when the ethylene is produced by cracking, the total energy loss exceeds 45% [3]. Along with China's rapid development, the total ethylene production capacity in 2017 increased to 23.21 million tons and the ethylene equivalent consumption grew up to 10% [4].

However, the fast growth of the ethylene production leads to the increase of the energy consumption and carbon emissions, and reduction of the ethylene production efficiency. Therefore, how to increase the ethylene production efficiency and reduce the carbon emissions has become a problem in the world. Nowadays, with the high-speed development of artificial intelligence, more and more energy optimization and analysis models have been used to raise productivity and energy efficiency in complex petrochemical processes. Geng et al. developed an artificial neural network (ANN) based on self-organizing cosine similarity, which overcame shortcomings of the single-hidden layer network in building the ethylene production prediction model [5]. An index decomposition analysis (IDA) method, combined with an ANN and a data envelopment analysis (DEA) was proposed by Olanrewaju et al. to assess energy consumption [6] and the energy capacity of industry in South African [7]. However, the process of extracting features in experiments is very complicated and the above models are not stable for fluctuations in complex industrial production. Considering the characteristics of high data dimension and

noise in the industrial data, an extreme learning machines (ELM) method based on the IDA for energy optimization and analysis was proposed by Geng et al. [8]. However, these methods do not take local minimization problems and the convergence rate in traditional ANNs into account. Considering the shortcomings and the limitations of the ANNs, a convolutional neural network (CNN) based on the cross-feature (CF) is proposed. The CF can combine the correlation between features to obtain the input of the CNN, which can avoid over-fitting problems caused by fewer features. Then the CNN is designed as a three-layer structure and the Rectified Linear Unit (ReLU) is introduced to achieve better generalization capability and stability with boiler fluctuations in petrochemical industry. The developed model has higher accuracy and scalability compared with back-propagation (BP) neural network and the ELM on the basis of the University of California Irvine (UCI) benchmark datasets. Finally, applied in the ethylene production process, the proposed method can accurately predict the ethylene production with an average relative generalization error of 2.86%. As the number of ethylene production features is small, the CNN is prone to over-fitting. The CF can not only obtain an expanded feature space serving as the input of the CNN, but also fully explore the correlation among crude oils, streams, fuel, water and electricity, which are essential inputs of ethylene production plants. By analyzing the model prediction results and the actual ethylene production, the experimental results show that ethylene plants have greater energy efficiency improvement space and energy saving potential. The proposed method serves as guidance for the ethylene production factory by allotting crude oils, streams, fuel, water and electricity in a reasonable way, with the energy efficiency increase of 6.38% and a carbon emissions decrease of 5.29%.

The rest of the paper is designed as follows. The current_research results in energy prediction and analysis in complex petrochemical processes are shown in Section 2. Section 3 illustrates the details of the CF-CNN. Comparison of verification results of the CF-CNN, the BP and the ELM based on the UCI standard dataset is conducted in Section 4. We present a

case study of energy optimization and prediction for an ethylene production process based on the CF-CNN in Section 5. In the end, the summary is presented in Section 6.

## 2. Related work

With the high-speed development of the computer technology, more and more simulation and energy efficiency prediction model of industrial production processes has been built by using the information technology, including the complex petrochemical production processes. Cai et al. developed a novel method combining the principal component analysis (PCA) with the complex network theory in process monitoring of the chemical industry [9]. A novel method based on the PCA was developed for energy efficiency evaluation by Zhu et al. [10]. However, finding a suitable mapping space in the PCA to accommodate the training data is complicated. Middleton et al. proposed a novel linearized approach to optimize energy network models [11]. Gong et al. developed an ethylene production system combining the DEA with factor analysis for energy efficiency evaluation [12]. Han et al. developed a modified DEA cross model for complex industrial processes [13,14]. However, the above statistical methods based on feature engineering in specific fields cannot be transferred to other industrial production fields and have limited generalization ability. Therefore, the ANN is proposed. The ANN is proved to be an effective algorithm for dealing with recognition, regression and classification tasks [15]. Hinton et al. proposed the BP neural network with a gradient-based back propagation algorithm [16], one of the most common neural networks. Ni et al. estimated the spatial distribution of PM2.5 concentrations in the Beijing-Tianjin-Hebei region from 2014 to 2016 by using the BP neural network [17]. However, the convergence speed of the BP neural network is too slow and may fall into local optimum, which makes industrial modeling unstable. The ELM [18, 19] greatly enriched the ANN, compared with the BP neural network. Unlike other feedforward neural networks, the ELM randomly initializes the network weight of the single hidden layer. The outputs are calculated by the Moore-Penrose inverse matrix. Therefore, the speed of calculating outputs and the generalization precision are pretty high, and not easy to

fall into local optimum [20]. Han at al. proposed the predictive modeling method combining the ELM with affinity propagation [21]. Cao et al. developed an improved adaptive evolution ELM algorithm to optimize the learning network weights of hidden neurons [22]. Suresh et al. proposed a genetic algorithm based on real-code to select the initialization parameters in the ELM. This method also speeded up the process of calculating the optimal parameters of the network, thereby improving the performance of the ELM [23]. Liu et al. combined four hybrid models based on four signal decomposition algorithms to calculate the output weights of the ELM, and the model obtained good performance in wind speed prediction [24]. On the other hand, the single-hidden layer feedforward neural network has significant limitations for production prediction modeling based on complex petrochemical data.

Recently, deep learning has set off a research boom and gained a large number of followers. Deep learning explores complex structures in large datasets by using BP algorithms to change the parameters of the training model (for calculating errors in each layer from errors in the previous layer). CNNs have achieved great progress in processing images, video and voice [25]. The CNN is a biologically inspired, trainable architecture that learns characteristics. The CNN is consisted of a filter bank, some non-linear feature pooling layers. Through these stages, the CNN can construct a multi-level functional hierarchy [26]. Lee et al. [27] developed a CNN model, named FDC-CNN for fault detection and classification (FDC) in semiconductor manufacturing processes. Meanwhile, CNNs have inherent advantages for images and strong adaptability in the image processing field [28]. Malek et al. obtained a one-dimensional (1D) CNN model for making chemometric analyses from spectroscopic data by particle swarm optimization training [29]. Abdeljaber et al. merged feature extraction and classification blocks into a compact learning body, and constructed a 1D-CNN model for damage detection [30].

In addition, the output of a chemical process is determined by one or more factors, and some factors are sometimes decisive. Therefore, how to fully consider the internal correlation and interaction between these factors is the modeling basis of the complex petrochemical

industry modeling. The CF is one of the methods in feature engineering [31]. The CF fully considered the internal correlation between features [32], exploring the correlation between each feature and all other features [33, 34]. Jian et al. proposed the KL transform theory for feature intersection in the feature space [35]. Then the combined feature is expressed as a complex vector [36]. Through the above discussion, the CF-CNN is proposed to build the energy optimization and prediction model of complex petrochemical processes.

## 3. Convolutional neural network integrating cross-feature

In order to make better use of the internal correlation between the influencing factors, the CF is introduced into the CNN. The feature space obtained by the CF is used as the input of the CNN to establish the energy optimization and prediction model of complex petrochemical industrial processes.

### 3.1 One-dimensional convolutional neural network

The CNN is mainly composed of some convolution layers, pooling layers and a fully connected layer [37]. The typical structure of the 1D-CNN is shown in Fig. 1. It is important that the CNN makes full use of natural signal properties like local connections and shared weights, which reduces the dimension of features and keeps an invariance of few shifts and distortions [38]. The back propagation and the forward propagation in the training process in 1D-CNN can be depicted as the following.

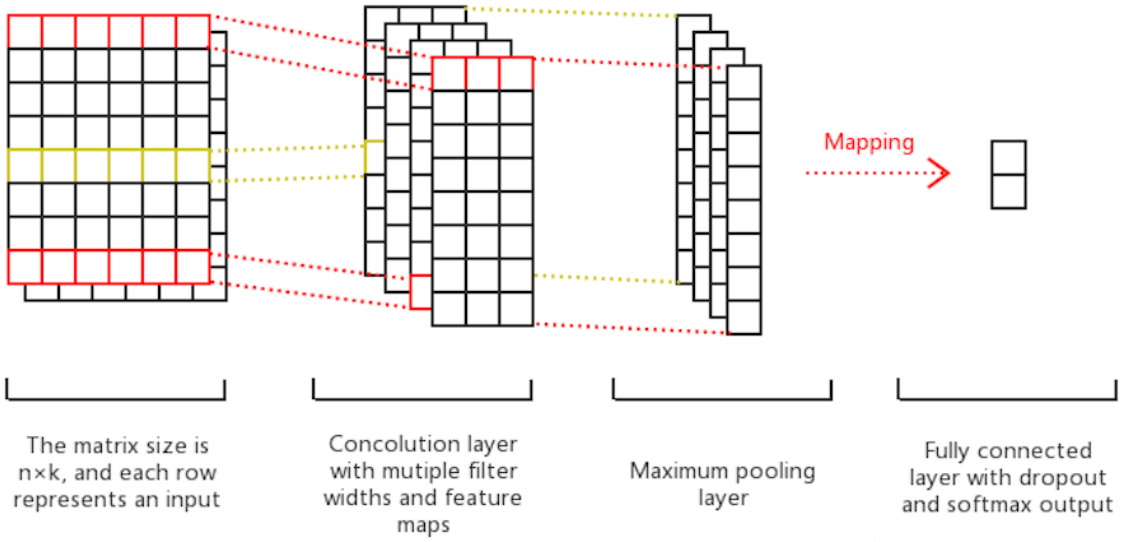


Fig. 1 The typical structure of the 1D-CNN

The convolutional layer can explore local connections of the previous layer of the network, and in the convolutional layer local units share the weight of the convolution kernel [39]. And the calculation formula for 1D-convolution is explained as follows:

$$y(k) = h(k) * u(k) = \sum_{i=0}^{N} h(k-i)u(i) \quad (1)$$

wherein $h(k)$ and $u(k)$ are 1D-sequences and $y(k)$ is the convolution result. $h(k) * u(k) = u(k) * h(k)$, which means the order can be swapped in the convolution calculation.

It can be seen from Fig. 1 that $H$ is a complex vector into which the input data is converted, and $C$ is the convolution kernel. Therefore, $P$ can be obtained after the convolution operation.

$$P = H * C \quad (2)$$

wherein $H$,$C$ can be written as:

$$H = [H_1, H_2, \ldots, H_n]^T = \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nk} \end{bmatrix}_{n \times k} \quad (3)$$

$$C = [C_1^T, C_2^T, \ldots, C_s^T]_{m \times s} \ (m < k) \quad (4)$$

wherein, $H_i(i = 1,2,\ldots,n)$ is a row vector which is represented as $[h_{i1}, h_{i2}, \ldots, h_{ik}]$ . $C_i^T$ (i = 1,2,…,s) is a column vector of the i-th convolution kernel. $n$ is the amount of the sample data, and $k$ represents the number of input features. $m$ and $s$ represent the length and the number of

the convolution kernel, respectively. Based on the above representations, Eq. (2) can be transformed to Eqs. (5) and (6).

$$P = [P_1, P_2, \ldots, P_s] \tag{5}$$

$$P_j = \begin{bmatrix} H_1 * C_j^T \\ \vdots \\ H_n * C_j^T \end{bmatrix}_{n \times (k-m+1)} \quad (j = 1,2, \ldots, s) \tag{6}$$

wherein $P_j$ represents the result of the convolution operation of the input vector $H$ and the j-th convolution kernel. The width of $P_j$ is $k - m + 1$. And a convolution operation with an input with k features and a convolution kernel of length m (assuming that m=3) is shown in Fig. 2.

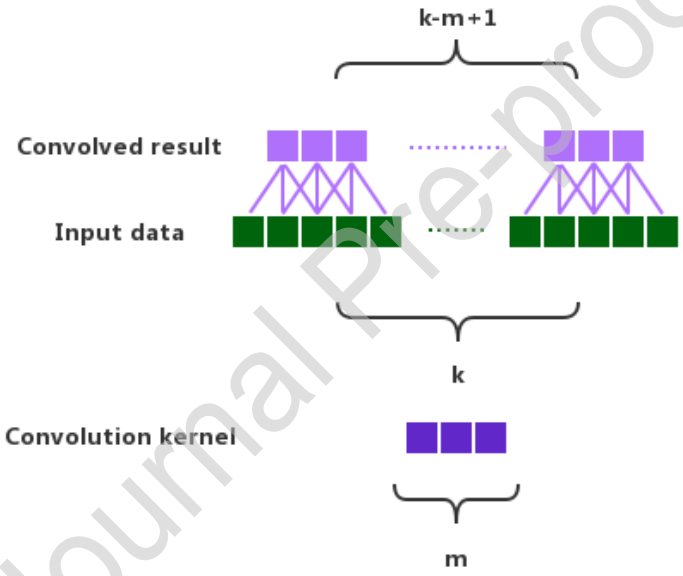


Fig. 2 1D convolution operation

In addition, the convolved result is subjected to a nonlinear mapping of the activation function, which makes the network converge more quickly [40]. There are many types of activation functions and the Rectified Linear Unit (ReLU) is used here, denoted by $G(\cdot)$ as follows:

$$P = G(P) = [G(P_1), G(P_2), \ldots, G(P_s)] \tag{7}$$

where $G(\cdot)$ is defined as follows:

$$G(x) = \begin{cases} x, & x > 0 \\ 0, & x \le 0 \end{cases} \tag{8}$$

The pooling layer can combine features with similar semantics to achieve nonlinearity [41]. The most common pooling layers include the maximum pooling and the average pooling, while the former one is used in this paper. The maximum pooling operation with the pooling layer of the size a=2 expressed as $max_{\mathrm{p}}$ is described in Fig. 3. The $P$ obtained before is calculated as in Eqs. (9) and (10).

$$J = maxp(P) = [J_1 J_2 ... J_s] \tag{9}$$

$$J_i = \begin{bmatrix} j_{11} & \cdots & j_{1(k-m-a+2)} \\ \vdots & \ddots & \vdots \\ j_{n1} & \cdots & j_{n(k-m-a+2)} \end{bmatrix}_{n \times (k-m-a+2)} \quad (i = 1,2,...,s) \tag{10}$$

More combinations of convolution and pooling are stacked to go through the fully connected layer for classification or regression purposes. The BP gradients in convolutional networks, like in normal networks, allow training of weights in all filters.

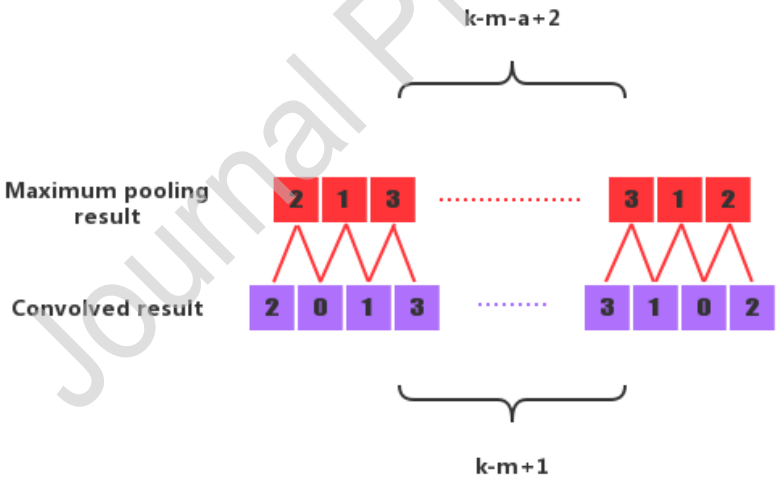


Fig. 3 The maximum pooling operation

### 3.2 Cross-feature

The CF is one important method in feature engineering. There are three main schemes including a composite feature vector, shorter feature vectors obtained through feature selection and conversion techniques, and multiple classifiers, all trained on their own feature sets.

Addition, subtraction, multiplication and division can be used to obtain new feature space [34]. Therefore, a new feature space can be constructed by combining original features.

Then the new feature set is generated by Cartesian product operation. Supposing the feature set is $A = \{a_n | n = 1,2,...,N; a_n \in R\}$ with $N$ features. A feature set $B$ is constructed as $B = A$. The Cartesian product of feature sets $A$ and $B$ forms a new feature set $F$.

$$F = \{a_i b_j | a_i \in A \wedge b_j \in B\} \tag{11}$$

When the situation of $a_i b_j = a_j b_i$ in the $F$ feature set is removed, the feature set $F$ is represented as follows:

$$F = \left\{f_n | n = 1,2,...,\frac{N(N-1)}{2}; f_n \in R\right\} (N \geq 3) \tag{12}$$

wherein the limit $N \geq 3$ is to make the number of features in the feature set $F$ greater than or equal to the original feature set $A$.

### 3.3 Evaluation criterion

In order to measure the validity of the developed model, the average relative generalization error (ARGE) is used to calculate errors between the training data and the test data [42], which can be calculated by Eqs. (13) and (14).

$$RGE_i = Abs\left(\frac{ModelOut_i^{inver} - ExpectOut_i^{inver}}{ExpectOut_i^{inver}}\right) \tag{13}$$

$$ARGE = \frac{\sum_{i=1}^{n} RGE_i}{n} \tag{14}$$

wherein $ModelOut_i^{inver}$ and $ExpectOut_i^{inver} (i = 1,2,...,n)$ are the inverse normalized prediction value and the real output of the train and test data. Since the data are normalized to the [0, 1] interval, it is essential to convert the normalized value back to the source data during the evaluation. And $n$ is the number of samples.

### 3.4 The framework of the CF-CNN

The CF fully considers the correlation between features based on the original feature set, and generates a new feature set based on Cartesian product, and the new features are set as inputs of the CNN. Given a training set $\{(X_n,Y_n)|n=1,2,...,N;X_n\in R^{w1};Y_n\in R^{w2}\}$, where $X_n$ is the input complex vector and $Y_n$ is the expected value of each data sample. There are $N$ samples in the data set. The details of the CF-CNN are shown as the following.

Step 1: The input data $X_n$ and $Y_n$ are normalized to the [0,1] interval by the min-max normalization, which can be seen in Eq. (15).

$$x_k^{(m)}=\begin{cases}\frac{x_k-\min\limits_{1\le j\le n}\{x_j\}}{\max\limits_{1\le j\le n}\{x_j\}-\min\limits_{1\le j\le n}\{x_j\}}, & if\ \max\limits_{1\le j\le n}\{x_j\}\neq\min\limits_{1\le j\le n}\{x_j\}\\ -1, & other\end{cases} \tag{15}$$

Step 2: Then the input matrix of the network model is constructed as the form of $[B,L,D]$, where $B$ is the batch quantity, $L$ the batch size and the length of the sequence, and $D$ the number of features. The dimension of the input data is shown in $\begin{bmatrix}x_1^{(1)} & \cdots & x_1^{(D)}\\ \vdots & \ddots & \vdots\\ x_L^{(1)} & \cdots & x_L^{(D)}\end{bmatrix}$. The training step of the CF-CNN starts from zero, at which time the data of the first batch $X_0$ is entered for training. As the iteration continues, when the counting step reaches $B-1$, the last batch data is entered into the CF-CNN. All the data are trained once.

Step 3: Initialize the parameters of the CF-CNN.

Step 4: According to Eqs. (2)-(12), parameters in the CF-CNN are continuously calculated and updated, retaining the best model in the training process. Here two sets of convolutional and maximum pooling layers are used, and the last fully connected layer integrates learned features for prediction.

Step 5: The actual output of the CF-CNN needs to be inversely normalized based on Eq. (16).

$$y_k=y_k^{(m)}*\left(\max\limits_{1\le j\le n}\{x_j\}-\min\limits_{1\le j\le n}\{x_j\}\right)+\min\limits_{1\le j\le n}\{x_j\} \tag{16}$$

where $y_k^{(m)}$ is the output of the CF-CNN, valued between [0,1].

Step 6: Compare the expected output with the output of the CF-CNN, then calculate the $ARGE$ between training data and test data according to Eq. (14).

The flow diagram of the CF-CNN is shown in Fig. 4.

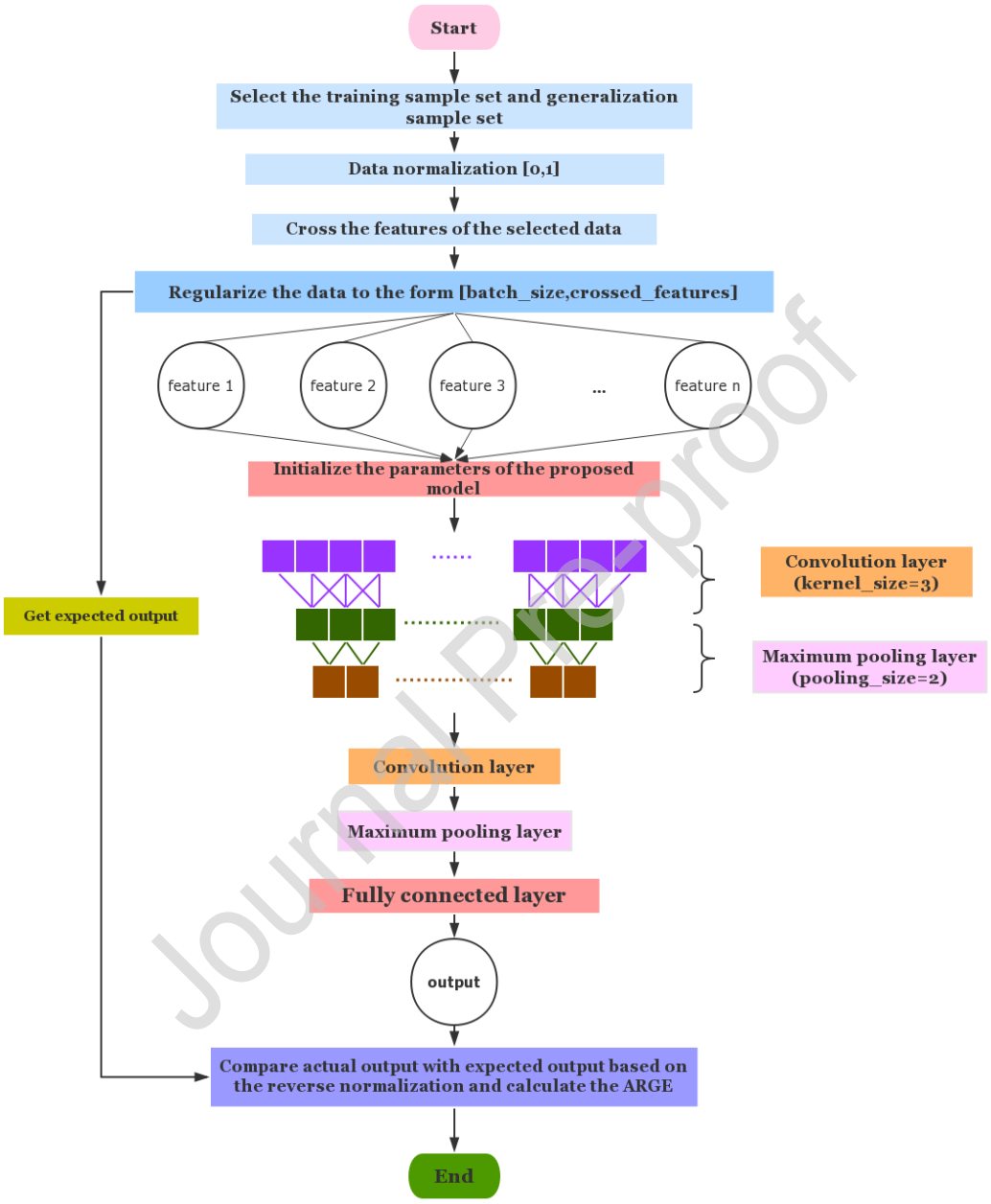


Fig.4 The flow diagram of the CF-CNN

## 4. Benchmark verification of UCI data sets

For the sake of proving the stability and capability of the proposed CF-CNN, the Combined Cycle Power Plant (CCPP) and the Airfoil Self-Noise data sets from UCI standard dataset

(more details can be found in the following website: http://archive.ics.uci.edu/ml/datasets.html) are picked to solve other regression problems. The characterization of the data set is presented in Table 1.

Table 1 The characterization of UCI data

| Data Set | Input Size | Output Size | Train Samples | Test Samples |
|---|---|---|---|---|
| Airfoil | 5 | 1 | 1000 | 147 |
| CCPP | 4 | 1 | 200 | 34 |

First, the feature space of the two data sets is processed, and the CF is used to construct a new feature space based on the original feature set. According to Eqs. (11) and (12), the number of new features in the two data sets is 10 and 6 respectively. Then the test operation is performed according to the modeling process described above. At the same time, in order to meet conditions of the comparative experiment, the CF is introduced into the BP (CF-BP) and ELM (CF-ELM) respectively. Then the CF-ELM and CF-BP are used to test the above data and compared with the CF-CNN in test results. The learning rate of the CF-BP is set as 0.01. Meanwhile, the training times are 5000 and the momentum factor is 0.9. The evaluation indexes of these networks are calculated based on the ARGE. The number of nodes in the hidden layer is adjusted according to the performance of these models, in order to achieve the best performance.

The test results of the two UCI data sets of three models are described in Table 2 and Table 3, respectively.

As can be seen from Tables 2 and 3, the CF-CNN can make the evaluation index reach the minimum value with the best performance. Meanwhile, the proposed model has better generalization accuracy and better performance compared with the CF-BP and the CF-ELM. Therefore, the CF-CNN is proved effective.

Table 2 The test results of Airfoil Data

| Model | Hidden nodes | ARGE |
|---|---|---|
| The CF-BP | 60 | 0.202 |
| The CF-ELM | 87 | 0.113 |
| The CF-CNN | - | 0.098 |

Table 3 The test results of CCPP Data

| Model | Hidden nodes | ARGE |
|---|---|---|
| The CF-BP | 50 | 0.268 |
| The CF-ELM | 56 | 0.178 |
| The CF-CNN | - | 0.164 |

## 5. Case study: Energy optimization and prediction modeling of petrochemical industries

For the purpose of verifying the effectiveness of the proposed model in the actual petrochemical production process, an energy optimization and prediction model is established based on the CF-CNN to explore the rationality of energy efficiency and carbon emissions.

### 5.1 Introduction of data

The ethylene industry is an important branch of the complex petrochemical industry. There are mainly two parts in the ethylene production: separation and cracking, in which a large amount of fuel is consumed during the cracking operation. At the same time, the fuel consumed reduces the energy efficiency. The separation process can be broken down into three steps: rapid cooling, compression and separation [43].

Different factories use different energy division standards and methods for calculating energy efficiency. This paper refers to the ethylene trade standard GB/T 2589-2008 and DB 37/751e2007 [44, 45]. The ethylene plants in the actual ethylene production process are divided into different technology indices and scale indices, and we calculate the energy consumption

index according to a unified calculation standard [8]. Fig. 5 shows a schematic flow diagram of the ethylene plant.

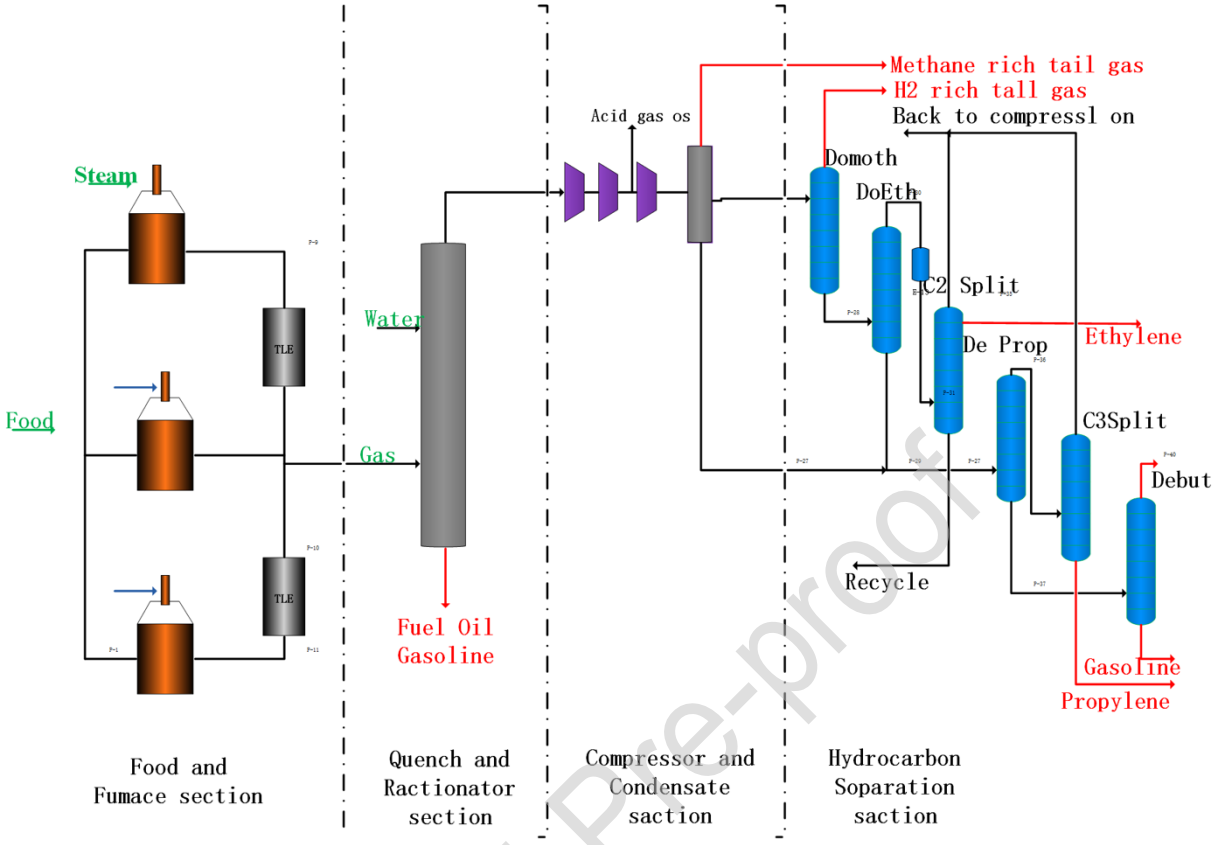


Fig. 5 Schematic flow diagram of the ethylene plant

There are many factors influencing the energy efficiency of the ethylene industry, and the most important factors are raw materials, fuel, power and ethylene production. Raw materials have a variety of crude oil, including hydrogenation tail oil, C345, light diesel fuel, naphtha and others. The fuel includes fuel gas, light weight oil and heavy weight oil. Power includes the electricity, and water includes boiler water and other water. The steam consists of super high pressure steam, high pressure steam, medium pressure steam and low pressure steam. The production indicator is expressed as the sum of ethylene, propylene and C4. According to the conversion relationship in the energy calculation method [46], the measurement units of fuel, steam, water and electricity are uniformly converted into GJ/Ton ethylene. However, the crude oil and the production indicator are still measured in Ton.

The data source comes from twenty ethylene plants based on different technologies in China [8]. The monthly industrial data dating from 2014 to 2018 for 5 years are used to realize the energy optimization and production prediction of the ethylene production using the proposed model. The input variables of the proposed model are crude oil, fuel, steam, water and electricity, and the output variable is the ethylene production, as described in Fig. 6. In the Fig. 6, the X-coordinate axis represents the input variables and the output variable with their own units, the Y-coordinate axis represents all the samples, and the Z-coordinate axis represents the values of variables.

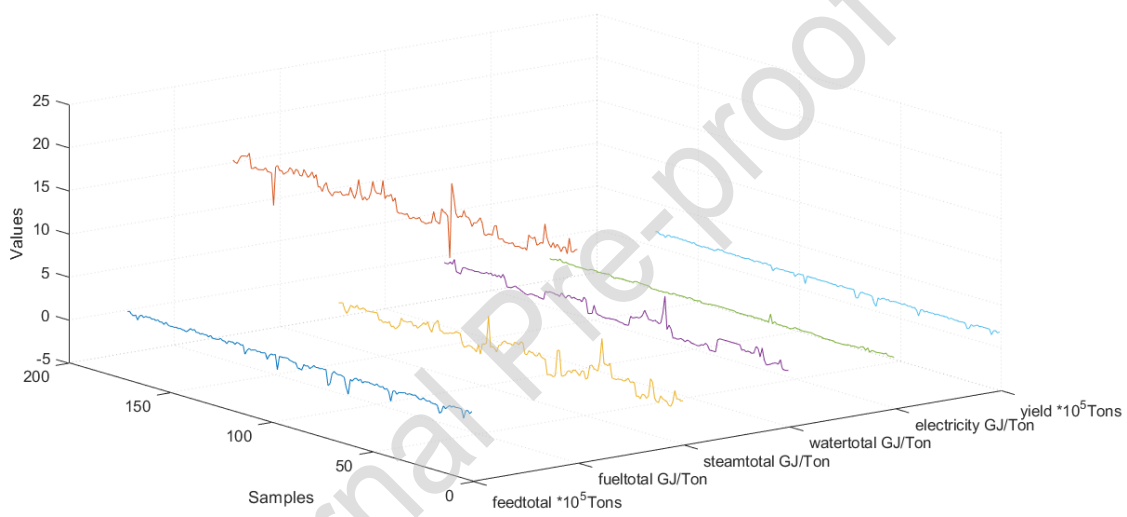


Fig. 6 The input variables and the ethylene yield

In order to ensure the data quality, the extracted data is preprocessed, including data cleaning and data reduction. During the data cleaning phase, samples with outliers of the input or output variables are excluded to ensure that all samples are valid. Then, the input features are mapped into a new feature space according to Eqs. (11) and (12). The data is normalized according to Eq. (15).

**5.2 Energy optimization and prediction modeling of ethylene plants**

The preprocessed data is split into training data and generalization data. The proposed CF-CNN are compared with the CF-ELM and the CF-BP to prove the predictive performance of

the proposed model. Initial values of the network parameters of three prediction models are shown in Table 4. The $ARGE$ is used to estimate the prediction accuracy in the experiment. The nodes of the hidden layer and the prediction accuracy of the training and test sets are shown in Table 5. The relationship between the predictive performance of compared networks and the number of hidden layer nodes is depicted in Fig. 7.

Table 4 Initial values of the network parameters

| Model parameters | The CF-BP | The CF-ELM | The CF-CNN |
|---|---|---|---|
| Input layer node number | 5 | 5 | 5 |
| Input node number after CF | 10 | 10 | 10 |
| Output layer node number | 1 | 1 | 1 |
| Number of iterations | 5000 | - | 200 |
| Activation function | sigmoid | - | ReLU |

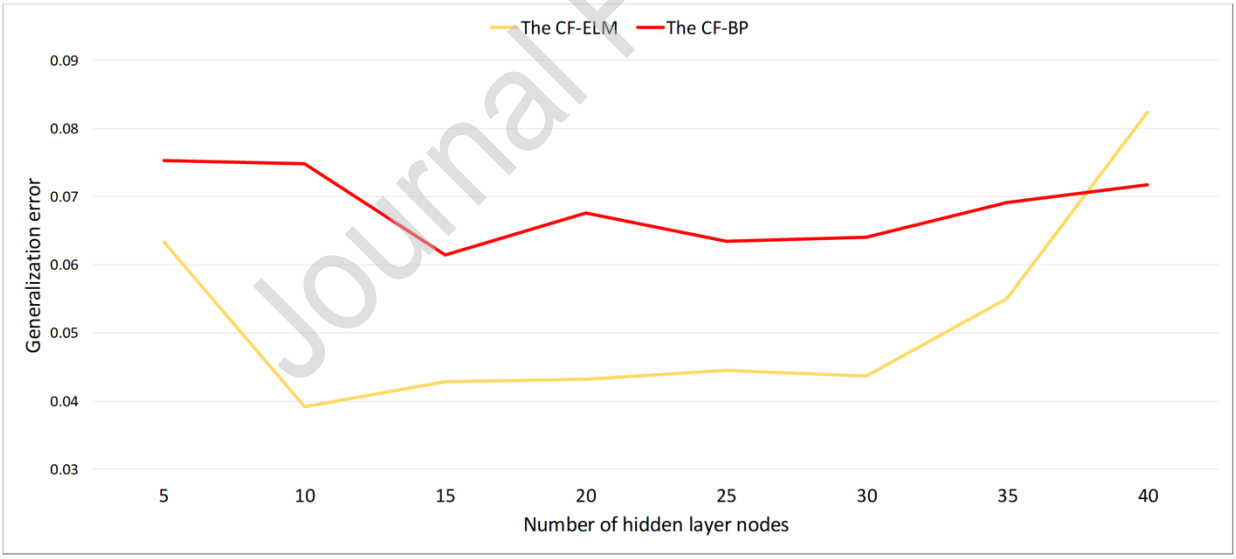


Fig. 7 The relationship between the number of hidden layer nodes and the generalization error of the CF-BP and the CF-ELM

Table 5 The result of different models

| Ethylene Production | The CF-BP | The CF-ELM | The CF-CNN |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Hidden layer node | 15 | 10 | - |
| Training ARGE | 0.0592 | 0.0382 | 0.023 |
| Generalized ARGE | 0.0615 | 0.0391 | 0.0286 |

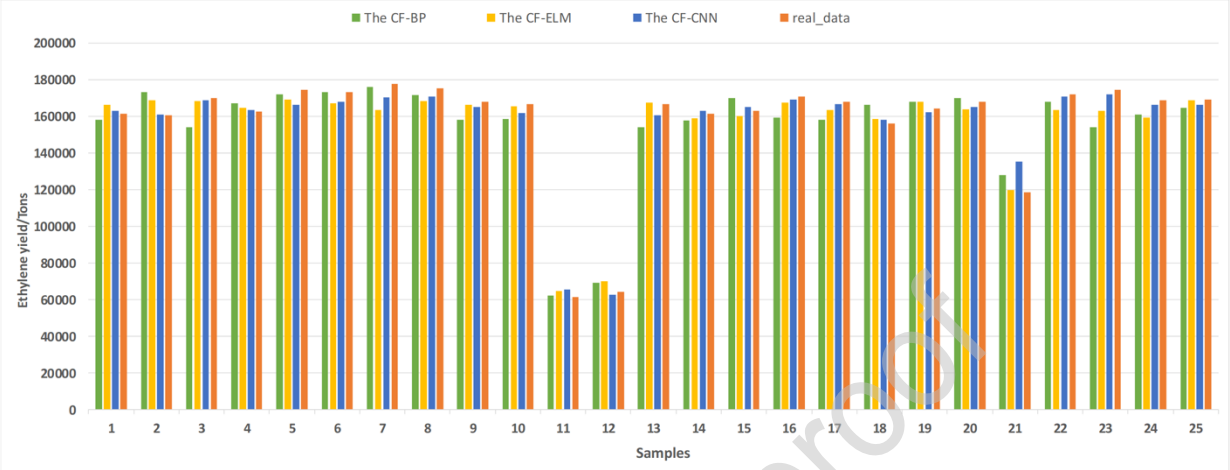


Fig. 8 Predicted values by different models and actual values of ethylene production plants

As can be seen from Fig. 7, when the number of hidden layer nodes reaches 10 and 15, the CF-ELM and the CF-BP models get the minimum generalization error. Therefore, the CF-ELM and the CF-BP models that achieve the best performance are compared with the proposed model. As is evident in Table 5, the CF-CNN has the minimum error, which means that the CF-CNN is effective in improving the ethylene production efficiency. The ARGE of the CF-BP, the CF-ELM and the CF-CNN is 6.15%, 3.91% and 2.86%, respectively. The predicted ethylene production of three models for the test sample and the actual values of ethylene production plants are depicted in Fig. 8, which also shows that the CF-CNN is more capable and more stable to fit the data. Meanwhile the proposed CF-CNN can obtain more accurate prediction of the ethylene production, serving to achieve energy optimization.

The difference between predicted values of the CF-CNN model and actual values is shown in Fig. 9, which is expressed as $y_{real} - y_{pred}$. According to the data of samples 4, 11 and 21, the actual ethylene production amount is obviously lower than prediction outputs of the proposed model. For example, in the 21st group, the proposed model predicts the ethylene

production is 135349.34 tons, while the actual ethylene production amount is only 118750 tons, which indicates that the production efficiency of the ethylene plant is low in this month and the energy utilization rate is low. Therefore, it is necessary to adjust the structure of the production line and the proportion of energy and power inputs. However, according to data of the 3, 8, 19 and 20 groups, the actual ethylene production amounts are similar to the prediction outputs. The actual ethylene production amounts in the 5, 6, 10 and 13 groups are higher than the predicted values, and the energy utilization rate is quite high and needs to be maintained, which also indicates the direction of future improvement.

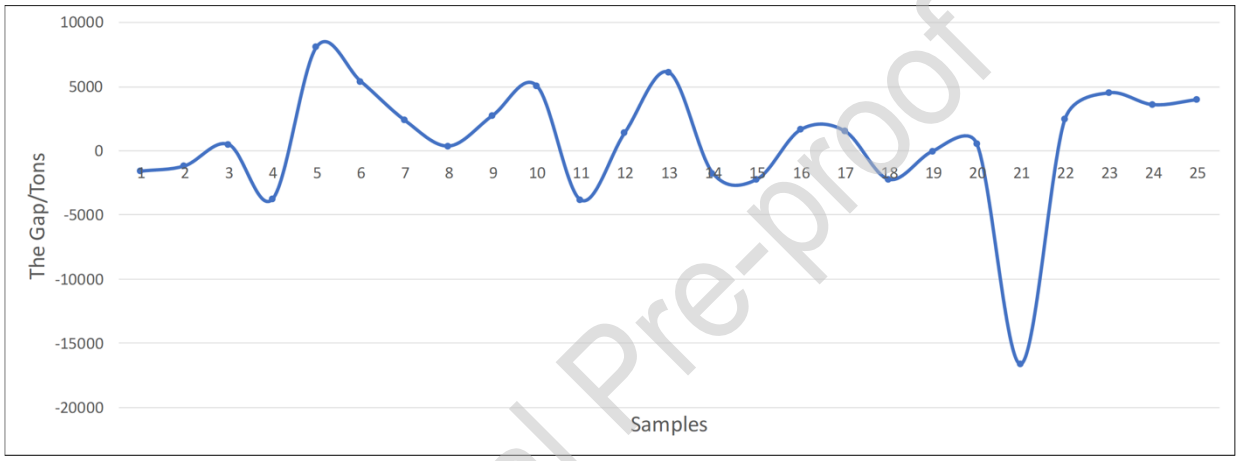


Fig. 9 The difference between the predicted result and the actual value

For example, in order to increase the ethylene production of the 11th test sample, the adjustment of each parameter in the production process is described in Fig. 10. According to data analysis of the forecast results, the crude oil, fuel, water and electricity in the 11th group of the ethylene plant needs to be decreased by 5,519 tons, 0.43GJ/Ton, 0.3724GJ/Ton and 0.105GJ/Ton, respectively. At the same time, the steam is increased by 1.223GJ/Ton, so that the production amount of ethylene increases by about 7,192 tons. In this way, the ethylene is increased from 0.564 tons to 0.6 tons per ton of crude oil, with an increase rate of 6.38%. In addition, 1.77 tons of crude oil per ton of ethylene is required before production adjustment. However, the oil is reduced to 1.66 tons after the adjustment. Meanwhile, the carbon emissions per ton of ethylene produced from crude oil are reduced from 1.48 tons to 1.39 tons, as the

carbon emission coefficient of crude oil is 0.8363 TC/Ton. Similarly, the carbon emissions per ton of ethylene produced from fuel are reduced from 0.392 tons to 0.383 tons according to the carbon emission coefficient of 21.1 KgC/GJ, with a decrease rate of 5.29% [47].

Considering the difference in the way the ethylene produced, the ethylene production of the 4th test sample is also analyzed, which is depicted in Fig. 11. The crude oil, fuel and electricity in the 4th group of the ethylene plant are decreased by 1424 tons, 1.45GJ/Ton and 0.003GJ/Ton, respectively. Meanwhile, the steam and water are increased by 0.134GJ/Ton and 0.139GJ/Ton, which can increase the ethylene production amount by 3851 tons. 0.574 tons of ethylene per ton of crude oil can be increased to 0.59 tons, with an increase rate of 2.89%. Furthermore, the carbon emissions per ton of ethylene produced decrease from 1.87 tons to 1.8 tons, with a decrease rate of 3.74%.

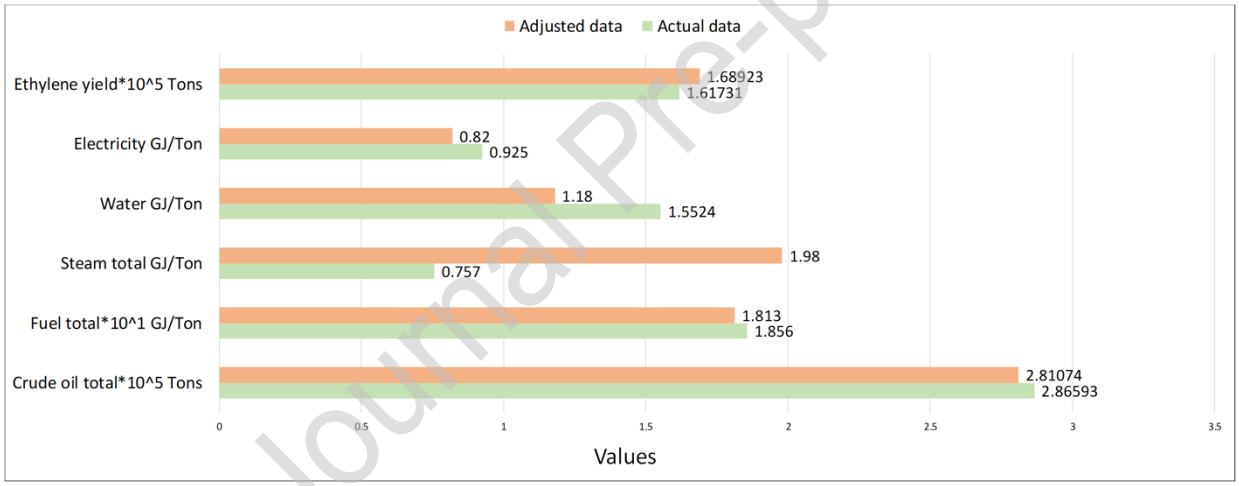


Fig. 10 The adjustment of parameters and increase of yield in ethylene production (11th)

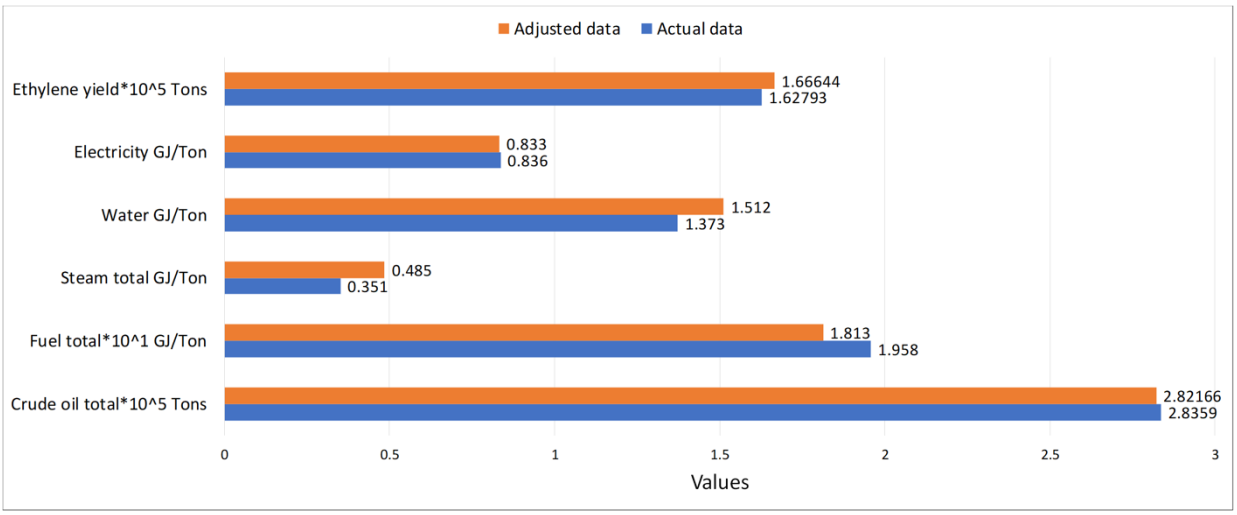


Fig. 11 The adjustment of parameters and increase of yield in ethylene production (4th)

## 6. Conclusion

The CF-CNN method is proposed for energy optimization and prediction modeling in complex petrochemical industries, which overcomes the shortcomings of the traditional ANN and the traditional statistical methods. The CF fully considers the nonlinear relationship and internal relationship between the features. Simultaneously, the new feature space generated from the original feature set is more conducive to improving the performance of the CF-CNN. In addition, the CF-CNN is examined by the UCI standard data sets with higher accuracy, better performance and greater generalization capabilities, compared with the CF-BP and the CF-ELM. Furthermore, the proposed model is applied in energy optimization and prediction modeling of ethylene plants in the complex petrochemical industry, with the ARGE of 2.86%. The model can accurately predict the ethylene production objectively and analyze the energy efficiency of different ethylene plants. By analyzing the model prediction results and actual ethylene production, ethylene plants have greater energy efficiency improvement space and energy saving potential. Furthermore, the CF-CNN guide the ethylene production factory by allotting crude oils, streams, fuel, water and electricity in a reasonable way with an energy efficiency increase of 6.38% and a carbon emissions decrease of 5.29%. This is significant for economic growth and environmental protection.

In the future, data with shorter time span will be used instead of monthly data to optimize our method, meeting with current production requirements. In addition, some other influencing factors will be added such as the cost that affects industrial profits. Since the proposed method requires a lot of data being continuously trained to obtain the best network weights, in the model needs long training time. Thus, parallel training methods will be introduced into our method to achieve optimization.

**Acknowledgments**

This work is partly funded by the National Natural Science Foundation of China (21978013, 61603025 and 61673046), Science and Technology Major Project of Guizhou Province (Guizhou Branch [2018]3002), Industrial Internet innovation and development project (TC19084DY) and the Fundamental Research Funds for the Central Universities (XK1802-4).

**References**

[1] Chen QQ, Lv M, Wang DF, Tang ZY, Wei W, Sun YH. Eco-efficiency assessment for global warming potential of ethylene production processes: A case study of China. Journal of Cleaner Production 2017,142(4):3109-3116.

[2] Chen JM, Yu BY, Wei YM. Energy technology roadmap for ethylene industry in China. Applied Energy 2018,224:160-174.

[3] Ghannadzadeh A, Sadeqzadeh M. Exergy analysis as a scoping tool for cleaner production of chemicals: a case study of an ethylene production process. Journal of Cleaner Production 2016,129:508-520.

[4] Luo HJ, Lv XD, Zhao R, Yang GY, Yang XX. 2017 review and 2018 outlook of global and China's petrochemical industry. International Petroleum Economics 2018,26(5):32-41.

[5] Geng ZQ, Li YN, Han YM, Zhu QX. A novel self-organizing cosine similarity learning network: An application to production prediction of petrochemical systems. Energy 2018,142:400-410.

[6] Olanrewaju OA, Jimoh AA, Kholopane PA. Integrated IDA-ANN-DEA for assessment and optimization of energy consumption in industrial sectors. Energy 2012,46(1):629-635.

[7] Olanrewaju OA, Jimoh AA, Kholopane PA. Assessing the energy potential in the South African industry: a combined IDA-ANN-DEA (index decomposition analysis-artificial neural network-data envelopment analysis) model. Energy 2013,63:225-232.

[8] Geng ZQ, Yang X, Han YM, Zhu QX. Energy optimization and analysis modeling based on extreme learning machine integrated index decomposition analysis: Application to complex chemical processes. Energy 2017,120:67-78.

[9] Cai E, Liu D, Liang L, Xu GH. Monitoring of chemical industrial processes using integrated complex network theory with PCA. Chemometrics and Intelligent Laboratory Systems 2015,140:22-35.

[10] Zhu L, Chen JH. Development of energy efficiency principal component analysis model for factor extraction and efficiency evaluation in large-scale chemical processes. International Journal of Energy Research 2019,43(2):814-828.

[11] Middleton RS. A new optimization approach to energy network modeling: anthropogenic CO2 capture coupled with enhanced oil recovery. International Journal of Energy Research 2013,37(14):1794-1810.

[12] Gong SX, Shao C, Zhu L. Energy efficiency evaluation based on DEA integrated factor analysis in ethylene production. Chinese Journal of Chemical Engineering 2017,25(6):793-799.

[13] Geng ZQ, Dong JG, Han YM, Zhu QX. Energy and environment efficiency analysis based on an improved environment DEA cross-model: Case study of complex chemical processes. Applied Energy 2017,205:465-476.

[14] Han YM, Long C, Geng ZQ, Zhu QX, Zhong YH. A novel DEACM integrating affinity propagation for performance evaluation and energy optimization modeling: Application to complex petrochemical industries. Energy Conversion and Management 2019,183:349-359.

[15] Liu Y, Xu LX, Li MZ. The Parallelization of Back Propagation Neural Network in MapReduce and Spark. International Journal of Parallel Programming 2017,45(4):760-779.

[16] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986,323:533-536.

[17] Ni XL, Cao CX, Zhou YK, Cui XH, Singh RP. Spatio-Temporal Pattern Estimation of PM2.5 in Beijing-Tianjin-Hebei Region Based on MODIS AOD and Meteorological Data Using the Back Propagation Neural Network. Atmosphere 2018,9(3):105.

[18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE International Joint Conference on Neural Networks 2004,2:985-990.

[19] Tang J, Deng C, Huang GB. Extreme learning machine for multilayer perceptron. IEEE Transactions on Neural Networks and Learning Systems 2017,27(4):809-821.

[20] Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. Neurocomputing 2006,70(1-3):489-501.

[21] Han YM, Wu H, Jia MH, Geng ZQ, Zhong YH. Production capacity analysis and energy optimization of complex petrochemical industries using novel extreme learning machine integrating affinity propagation. Energy Conversion and Management 2019,180:240-249.

[22] Cao JW, Lin ZP, Huang GB. Self-adaptive evolutionary extreme learning machine. Neural Processing Letters 2012,36(3):285-305.

[23] Suresh S, Saraswathi S, Sundararajan N. Performance enhancement of extreme learning machine for multi-category sparse data classification problems. Engineering Applications of Artificial Intelligence 2010,23(7):1149-1157.

[24] Liu H, Tian HQ, Li YF. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Conversion and Management 2015,100:16-22.

[25] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015,521:436-444.

[26] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional Networks and Applications in Vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems 2010:253-256.

[27] Lee KB, Cheon S, Chang OK. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. IEEE Transactions on Semiconductor Manufacturing 2017,30(2):135-142.

[28] Alex K, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 2012:1097-1105.

[29] Malek S, Melgani F, Bazi Y. One-dimensional convolutional neural networks for spectroscopic signal regression. Journal of Chemometrics 2017,32(5):2977

[30] Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M, Inman DJ. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. Journal of Sound and Vibration 2017,388:154-170.

[31] Turner CR, Fuggetta A, Lavazza L, Wolf AL. A conceptual basis for feature engineering. Journal of Systems and Software 1999,49(1):3-15.

[32] Alfred A, Nancy G. Feature Interactions in the Global Information Infrastructure (Panel). ACM Sigsoft Software Engineering Notes 1995.

[33] Huang YA, Fan W, Lee WK, Yu PS. Cross-Feature Analysis for Detecting Ad-Hoc Routing Anomalies. 23rd International Conference on Distributed Computing Systems 2003:478-487.

[34] Schuller B, Reiter S, Rigoll G. Evolutionary Feature Generation in Speech Emotion Recognition. 2006 IEEE International Conference on Multimedia and Expo:5-8.

[35] Yang J, Yang JY. Generalized K–L transform based combined feature extraction. Pattern Recognition 2002,35(1):295-297.

[36] Yang J, Yang JY, Zhang D, Lu JF. Feature fusion: parallel strategy vs. serial strategy. Pattern Recognition 2003,36(6):1369-1381.

[37] Christof A, Leopold PT, Oliver S. Deep learning for computational biology. Molecular Systems Biology 2016,12(7):878.

[38] Mallat S. Understanding deep convolutional networks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2016,374(2065).

[39] Guo YM, Liu Y, Oerlemans A, Lao SY, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 2016,187:7-48.

[40] Qian S, Liu H, Liu C, Wu S, Wong HS. Adaptive activation functions in convolutional neural networks. Neurocomputing 2018,272:204-212.

[41] Liu L, Shen C, Hengel AD. Cross-Convolutional-Layer Pooling for Image Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017,39(11):2305-2313.

[42] Jackowski T, Woźniakowski H. Complexity of approximation with relative error criterion in worst, average, and probabilistic settings. Journal of Complexity 1987,3(2):114-135.

[43] Han YM, Geng ZQ, Zhu QX, Qu YX. Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry. Energy 2015,83:685-695.

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

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

[46] Calculation method for energy consumption in petrochemical engineering design (SH/T3110-2001), 2002.

[47] Zhao M, Zhang WG, Yu LZ. Carbon Emissions from Energy Consumption in Shanghai City. Research of Environmental Sciences 2009,22(8):984-989.

Declaration of interest statement:

The authors declared that they have no conflicts of interest to this work.

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

We appreciate your warm and earnest work, and hope that the revision will meet with requirements.

Once again, thank you very much for your comments and suggestions.

Sincerely yours,

Zhiqiang Geng, Yanhui Zhang, Chengfei Li, Yongming Han, Yunfei Cui, Bin Yu

**Highlights**

- A novel convolutional neural network (CNN) integrating cross-feature is proposed.
- Energy optimization and prediction model of petrochemical industries is obtained.
- The energy saving potential of ethylene production plants is about 6.38%.
- Carbon emissions reduce by 6.08% in ethylene production plants.