Energy Conversion and Management 183 (2019) 349–359

Contents lists available at ScienceDirect

Energy Conversion and Management

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

A novel DEACM integrating affinity propagation for performance evaluation and energy optimization modeling: Application to complex petrochemical industries

Yongming Han[a,b], Chang Long[a,b], Zhiqiang Geng[a,b,*], Qunxiong Zhu[a,b,*], Yanhua Zhong[c]

[a] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China
[c] Jiangmen Polytechnic, Jiangmen, Guangdong 529020, China

ARTICLE INFO

Keywords:
Data Envelopment Analysis (DEA)
DEA cross-model (DEACM)
Affinity propagation (AP)
Performance evaluation
Energy optimization
Complex petrochemical industries

ABSTRACT

Data Envelopment Analysis (DEA) has been widely used in performance and energy efficiency evaluation. However, in the traditional DEA, the effective of each decision making unit (DMU) is evaluated through its own optimized perspective and regardless of other DMUs influence, which may result in too many effective DMUs. And the DEA cross-model (DEACM) can distinguish the effective DMUs better by constructing a cross-efficiency matrix, but the optimal weight of the DMU may not be unique, so the cross efficiency of the DEACM may be different. Therefore, this paper proposes a novel DEACM based on the affinity propagation (AP) clustering algorithm (AP-DEACM). Through the AP clustering algorithm, the high impact data affecting the performance capacity and energy saving are obtained. Then the better effective DMU is identified through the high discrimination of the AP–DEACM. Finally, the proposed AP-DEACM is used for performance evaluation and energy optimization modeling of the Pure Terephthalic Acid (PTA) production process and the ethylene industrial process in complex petrochemical industries. The experimental results show that the energy saving potential of PTA production plants and ethylene production plants are 2.78% and 1.26%, respectively, and the average value of carbon emission savings potential is 3.62% in ethylene production plants.

1. Introduction

Energy is the cornerstone of modern civilization, and the unsustainable energy production model has become an obstacle to social progress [1]. Therefore, improving the energy efficiency is considered to be an important means of achieving economic goals. The ethylene industrial production is the basis of the petrochemical industry, which is usually used as one of the main indicators to measure the industry level of a country [2]. However, due to the factors such as raw materials and processes, the cost of the ethylene production in North America was only \$350 to \$400/Ton from 2011 to 2014, while in China during the same period exceeded \$1100/Ton [3], which was much higher than the international advanced level. Among them, the energy consumption of the ethylene production process exceeds 50% of the operating cost of the plant [4]. The Pure Terephthalic Acid (PTA) is an important raw material for polyester products in the petrochemical industry [5]. The PTA production in China has soared from 3 million Tons in 2000 to 50 million Tons in 2016. And corporate profit margins of the PTA overcapacity are compressed [6]. Therefore, the performance evaluation and energy optimization of the PTA production process and the ethylene production process has high economic benefits [4].

If the model can be used to predict the production of the petrochemical industry by analyzing the production data, the goal of energy optimization can be achieved. Therefore, this paper proposes a performance evaluation and energy optimization model based on the improved DEACM integrated the affinity propagation (AP) clustering (AP-DEACM). The key factors from the raw multi-dimensional data of the production process can be obtained through the AP clustering. Then the clustering results are set as inputs of the AP-DEACM to analyze the production capacity and optimize the energy configuration. Finally, a performance capacity evaluation and energy optimization model for the ethylene production system and the PTA production system in the complex petrochemical industry is established based on the improved method. Compared with the general DEA model and the DEA cross-model, the validity and the practicability of the proposed method are verified by the experimental results. Moreover, the improved model can

* Corresponding authors at: College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China.
E-mail addresses: gengzhiqiang@mail.buct.edu.cn (Z. Geng), zhuqx@mail.buct.edu.cn (Q. Zhu).

https://doi.org/10.1016/j.enconman.2018.12.120
Received 20 October 2018; Accepted 31 December 2018

be used in performance evaluation and energy optimization of PTA production plants and ethylene production plants. The experimental results show that the energy saving potential of PTA production plants and ethylene production plants are 2.78% and 1.26%, respectively, and the average value of carbon emission savings potential is 3.62% in ethylene production plants.

The remaining chapters are arranged as follows: Section 2 introduces the relevant literature of the performance analysis and energy saving, and Section 3 describes the performance evaluation and energy optimization model of complex chemical processes based on the AP-DEACM. Section 4 presents two case studies of energy optimization and analysis in the ethylene production and PTA production industries based on the AP-DEACM. Finally, discuss and draw conclusions in Sections 5 and 6, respectively.

2. Related work

In previous studies, the energy efficiency was analyzed using the mean method and the exponential method [7]. However, these methods do not provide a good indication of the objective relationship between input and output metrics and energy consumption. Geng et al. proposed an energy efficiency analysis framework for multi-region comparison based on the index decomposition analysis (IDA) to rank the energy consumption of China ethylene plants [8]. Although the IDA method can quantify the energy performance indicators, it did not indicate the specific improvement and the improvement direction of inefficient production equipments. The Stochastic Frontier Analysis (SFA) method is a typical representation of the energy efficiency analysis parameter method. Fetanat et al. use the SFA to estimate the energy-saving potential distribution of the industrial sector [9]. However, when dealing with multiple output situations, the SFA is too complex to be used in the complex petrochemical industry.

Compared with the SFA method, the input and output indicators of the DEA method can be multi-dimensional without knowing the functional relationship between variables [10]. Since the DEA method has been widely applied in the relative efficiency evaluation of industry [11]. Hernándezsancho et al. determined the determinants of energy consumption in a Spanish wastewater treatment plant based on non-radial DEA [12]. Song et al. proposed Bootstrap for DEA modification values based on small sample data to evaluate the overall energy efficiency of the BRICS countries [13]. Bi et al. divided the production system of coal-fired power plants into production processes and pollutant emission reduction processes. As a link between the two processes, pollutant intermediates proposed a new two-stage network DEA model to evaluate the performance of coal-fired power generation in China [14]. Azadeh [15] and Wen [16] et al. introduced fuzzy mathematics and established fuzzy DEA model (FMP) to analyze DMU. Song et al. evaluated energy efficiency (EE) of 34 coal-fired power units in China based on non-parametric DEA method [17]. Hua et al. objectively evaluated the carbon dioxide emission performance of various provinces in China based on the three-stage DEA model to eliminate environmental factors and random errors [18]. Han et al. proposed a novel energy analysis framework for petrochemical processes based on the DEA coupling interpretation structure model (ISM), which can obtain the maximum efficiency index of each DMU, and find the main influencing factors of the energy consumption of the ethylene production plant [19]. Geng et al. used the IDA to obtain an energy-influence indicator, and an improved strategy for the ethylene production using the DEA based on relaxation variables [20].

However, the traditional DEA model is based on the maximization of the efficiency of its own DMU, and lacks consideration of other DMUs [21]. Therefore, Sexton proposed the concept of the DEACM, evaluating its own efficiency value under the premise of fully considering conditions (cross-evaluation) of other DMUs [22]. Chen et al. considered the game relationship of DEACM to evaluate the energy efficiency of various provinces in China [23]. Cook et al. pointed out that the cross efficiency of the DEACM is usually arbitrary. To overcome the influence of meaningful secondary objectives, the unit invariant multiplication DEA model is used to calculate the cross efficiency score [24]. Gao et al. established a novel cross efficiency evaluation method based on super efficiency data envelopment analysis (SE-DEA) for power system planning. By comparing with interval hierarchy method and fuzzy hierarchy method, the efficiency value distribution of the DEA-based cross-efficiency evaluation method is obtained better [25]. Han et al. proposed a fuzzy DEACM to guide the energy utilization efficiency in the ethylene production process [26]. Geng et al. analyzed the performance of 19 ethylene plants in China based on the improved Malaquist Productivity Index (MPI) integrating the improved DEACM considering input and output factors [27]. An et al. combined the analytic hierarchy process (AHP) and the DEA to obtain all possible cross efficiencies between DMUs [28]. Wu et al. proposed a DEA based on multi-attribute selection method that fully considers the competitive factors in multiple criteria decision making (MCDM) [29].

It can be seen that most of the existing methods focus on improving the accuracy of the cross matrix to eliminate the subjectivity of the cross matrix construction. However, for multi-input and multi-output systems, excessive or inappropriate evaluation indicators may reduce the evaluation efficiency [10]. Azadeh et al. constructed a comprehensive method based on the DEA, the principal component analysis (PCA) and the numerical taxonomy (NT) for the performance evaluation and energy optimization in energy intensive manufacturing industries [30]. Gong et al. used the k-means clustering algorithm to classify the working conditions of ethylene production data, and used the DEA to evaluate the performance of DMUs under different working conditions and guide the improvement of the ethylene production efficiency [31]. Rotem et al. built a model for fault detection of ethylene compressors based on PCA [32]. Wang et al. revealed the relationship between power consumption and carbon dioxide emissions based on Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [33]. Hu et al. established a model based on the improved fuzzy C-means (FCM) clustering method for predicting the optimal production conditions in the iron ore sintering process [34].

In the above methods, the clustering algorithm based on partitioning, K-means is taken as an example, which is sensitive to the initial clustering center and the abnormal point. The clustering algorithm based on density-based, DBSCAN is used as an example, which is less effective for data sets with the uneven density. The PCA is likely to ignore information that contributes to the difference in samples among the contributing principal components. Since there is no need to specify the number of clusters in advance, the AP clustering algorithm is considered to be a deterministic algorithm and has been applied in fault diagnosis [35], urban water supply network [36], the location of the nodes in the solar radiation monitoring station (MS) network [37] and the establishment of the thermal power unit flue gas content [38]. Therefore, this paper proposes a performance evaluation and energy optimization model based on the novel AP-DEACM to save the energy and reduce carbon emissions in complex petrochemical industries.

3. The AP-DEACM

To ensure the accuracy and the objectivity of the energy efficiency analysis in complex petrochemical industries, we integrated the AP into the DEACM method. The improved AP-DEACM method can filter the non-critical information or redundant information in complex petrochemical industries for more accurately estimating the energy efficiency.

3.1. Briefly review of the AP

The AP was first proposed by Frey and Dueck in 2007 [39]. There is no need to define the number of clusters and select the initial class center for the AP in advance. Compared with those traditional PCA and the K-means, which are sensitive to the initial class center selection and easy to fall into the local optimal clustering algorithm, the AP is a deterministic clustering algorithm, and the clustering group error is lower in the clustering process. When starting clustering, the AP treats all data points as cluster centers, and by iteratively to do "information transfer" between data points to continuously search for appropriate cluster centers and automatically identify the locations and the number of cluster centers from data points. The cluster center is found by maximizing the sum of similarities between the sample and the nearest cluster center. The Euclidean distance is chosen as a similarity criterion of the AP, that is, the similarity between any two points x_i and x_j is shown as follows:

$$s(i, j) = -(x_i - x_j)^2, i \neq j \tag{1}$$

where the value range of s is $(-\infty, 0]$. The greater the similarity value, the closer the distance between points. The AP algorithm can be expressed as the following process:

(1) Calculate the similarity between data objects, thereby constructing a similarity matrix S of N data, and setting a preference P. Generally, the value of P is the median of the input similarity.
(2) Initialize the responsibility matrix R with $N \times N$ and the availability matrix A with $N \times N$ of all values are 0.
(3) Update *R(i, k)* with the attraction degree iteration based on Eq. (2).

$$R_{t+1}(i, k) = \begin{cases} S(i, k) - \max\limits_{j \neq k}\{A_t(i, j) + R_t(i, j)\} i \neq k \\ S(i, k) - \max\limits_{j \neq k}\{S(i, j)\} i = k \end{cases} \tag{2}$$

(4) Update *A(i, k)* with the attribution degree iteration based on Eq. (3).

$$A_{t+1}(i, k) = \begin{cases} \min\left\{0, R_{t+1}(k, k) + \sum_{j \notin \{i,k\}} max\{0, R_{t+1}(j, k)\}\right\} i \neq k \\ \sum_{j \neq k} \max\{0, R_{t+1}(j, k)\} i = k \end{cases} \tag{3}$$

(5) When the condition $A(k, k) + R(k, k) > 0$ is satisfied, the data point k is selected as the cluster center.
(6) When the number of iterations exceeds the threshold or the cluster center does not change for a certain number of iterations, the iteration is terminated, and the next step is performed. Otherwise, the process returns to step (3).
(7) Assign another data point to the exemplars using the nearest rule, and each data point is set to the most similar exemplar.

In order to avoid the oscillation, a damping factor $\lambda \in (0, 1)$ is introduced during the information transmission. Each message is set to a weighted sum of λ times of the previous iteration and 1-λ times of the updated value, as shown in the following Eqs. (4) and (5).

$$R_{t+1}(i, k) = (1 - \lambda)\cdot R_{t+1}(i, k) + \lambda\cdot R_t(i, k) \tag{4}$$

$$A_{t+1}(i, k) = (1 - \lambda)\cdot A_{t+1}(i, k) + \lambda\cdot A_t(i, k) \tag{5}$$

where the larger the damping coefficient is, the better the effect of eliminating the oscillation is, but the convergence speed of the AP algorithm is lowered. In general, λ is set to 0.5 or 0.9 by default [40].

3.2. The DEACM

The DEA method is a very effective tool for evaluating the resource allocation efficiency of multi-input and multi-output DMUs. However, in practical problems, more DMUs can obtain the maximum efficiency value of 1, which cannot further distinguish the advantages or disadvantages. To this end, cross-evaluation mechanism was introduced. As an improved method, DEACM uses DEA as a sorting tool for multi-criteria DMU, which avoids the unreasonable and extreme weight distribution of traditional input and output indicators by traditional DEA, can effectively distinguish the merits of each DMU.

Assuming that there are n objects in the evaluation system, then any object i is DMU_i. The input factors and expected are:

$X = (X_1, X_2, \cdots, X_M)$, $X \in R_+^M$; there are M kinds of factor inputs.

$Y = (Y_1, Y_2, \cdots, Y_S)$, $X \in R_+^S$; there are S kinds of expected outputs.

Assuming that $w = [w_1, w_2, \cdots, w_M]$ and $v = [v_1, v_2, \cdots, v_s]$ are the weight vectors of inputs and outputs, respectively, then the ratio of total output to total output of DMU_i is

$$\begin{cases} max\frac{Y_i^T v}{X_i^T w} = E_i \\ s.\ t.\ \frac{Y_i^T v}{X_i^T w} \leq 1 (1 \leq j \leq n), w \geq 0, v \geq 0, X_i^T w = 1 \end{cases} \tag{6}$$

According to the Charnes-Cooper transformation, the Eq. (6) can be transformed into an equivalent linear programming problem.

$$\begin{cases} \max Y_i^T v = E_i \\ s.\ t.\ Y_i^T w \leq X_i^T w (1 \leq j \leq n), w \geq 0, v \geq 0, X_i^T w = 1 \end{cases} \tag{7}$$

The cross-evaluation value is obtained based on Sexton's definition of cross-evaluation of the DEA model.

$$E_{ij} = \frac{Y_j^T v_i^*}{X_j^T w_i^*} \tag{8}$$

The evaluation value of the j-th unit from the i-th unit is denoted as E_{ij}. The larger E_{ij}, the more favorable it is to DMU_j, and the more unfavorable it is to DMU_i. The Cross Efficient Matrix (CEM) can be constructed by the cross-evaluation index.

$$E = \begin{bmatrix} E_{11} & E_{12} & \cdots & E_{1n} \\ E_{21} & E_{22} & \cdots & E_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ E_{n1} & E_{n2} & \cdots & E_{nn} \end{bmatrix} \tag{9}$$

The elements on the diagonal are general evaluation indices, and the elements on the off-diagonal are cross-evaluation indices. Calculating the arithmetic mean of the column direction for each DMU, that is

$$\bar{E}_i = \frac{\sum_{k=1}^n E_{ki}}{n} \tag{10}$$

$\bar{E}_i$ is the cross-evaluation score under the traditional definition.

3.3. The framework of the AP-DEACM

The main steps of the AP-DEACM are shown as follows:

Step 1: Data preprocessing. The data is normalized according to the characteristics of the data, and an initial similarity matrix is obtained.
Step 2: Based on the AP method, a new data set with the higher similarity is obtained.
Step 3: The self-evaluation value of each DMU is calculated by using the clustering result as the input of DEACM.
Step 4: Though considering the impact of other DMUs, the cross-evaluation matrix is built by calculating cross-evaluation values.
Step 5: The cross-evaluation scores of each DMU are calculated through the cross-evaluation matrix to obtain the best production configuration.
Step 6: Analysis of results, the performance evaluation and energy optimization can be obtained.

The flow chart of the AP-DEACM is shown in Fig. 1.

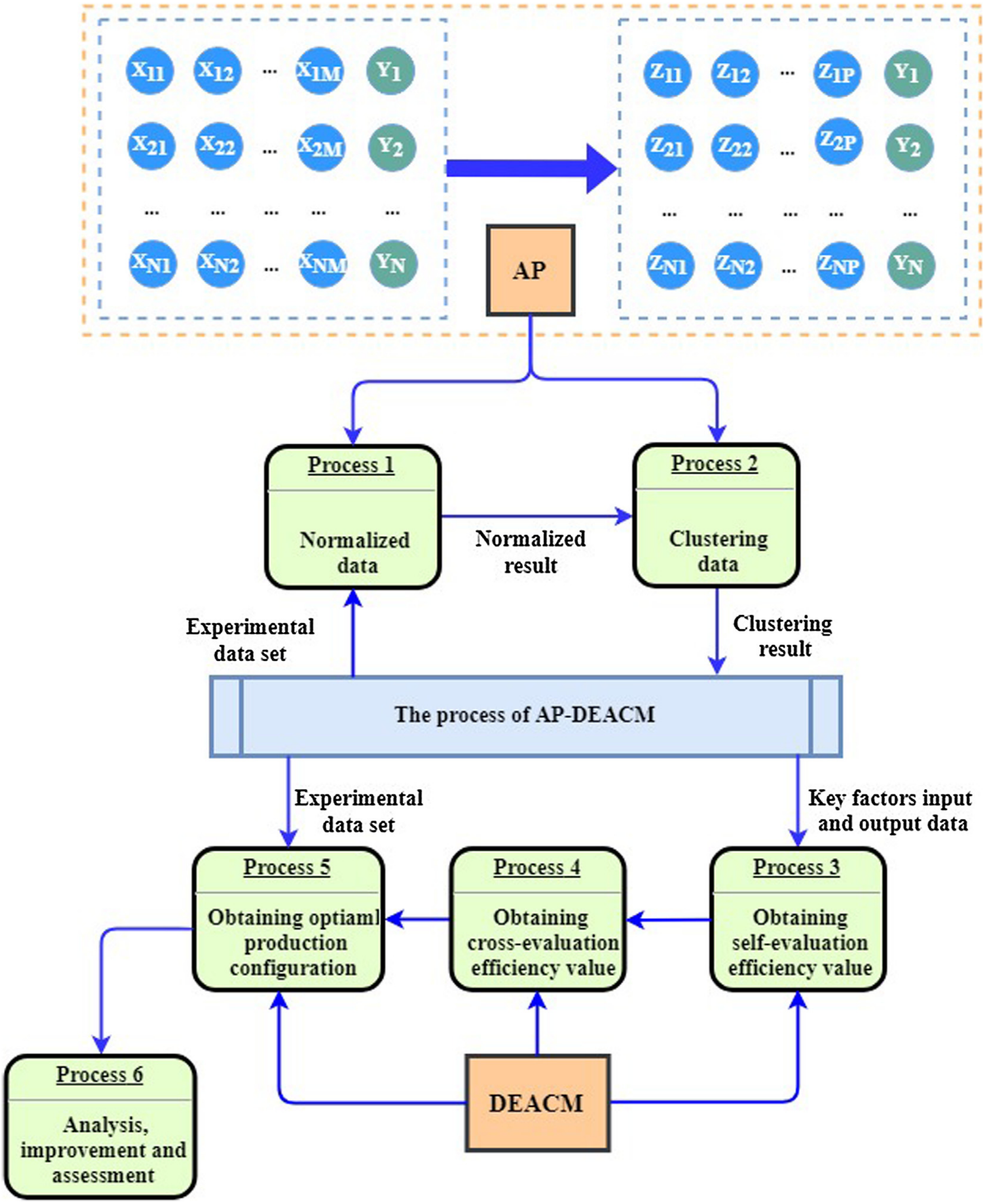

Fig. 1. The flow chart of the AP-DEACM.

4. Case Study: Performance evaluation and energy optimization model of complex petrochemical industries

In order to determine the availability and the effectiveness of the AP-DEACM in the petrochemical industry, the performance analysis and energy optimization model of PTA production processes and ethylene production processes is obtained based on the AP-DEACM. And the energy efficiency evaluation and energy optimization of PTA production process and ethylene production process have high economic benefits. Therefore, the PTA data comes from a petrochemical production enterprise [41], and the ethylene data comes from more than 20 ethylene production plants in China, including 7 production technologies [26].

4.1. Energy optimization for PTA production systems

In the PTA industry, the consumption of the acetic acid is the main indicator to measure the effectiveness and the advancement of the production plant. Therefore, one of the important ways to optimize the PTA production process is to reduce the consumption of acetic acid in the production process [42]. The PTA solvent system can be divided into three parts: solvent dehydration column, re-boiler and reflux tank [41]. The flow diagram of a solvent dehydrating tower of the PTA production plant is shown in Fig. 2.

By analyzing the operating characteristics of the solvent dehydration column, there are 17 factors including feed amount FC1501, temperature TI1504, reflux flow FC1502–FC1504, temperature TI15010, reboiler steam flow FC1507 and tower inside temperature TI1511–TI1519, TC1501 which affect the acetic acid content of the top of the tower. However, the actual acetic acid content is difficult to

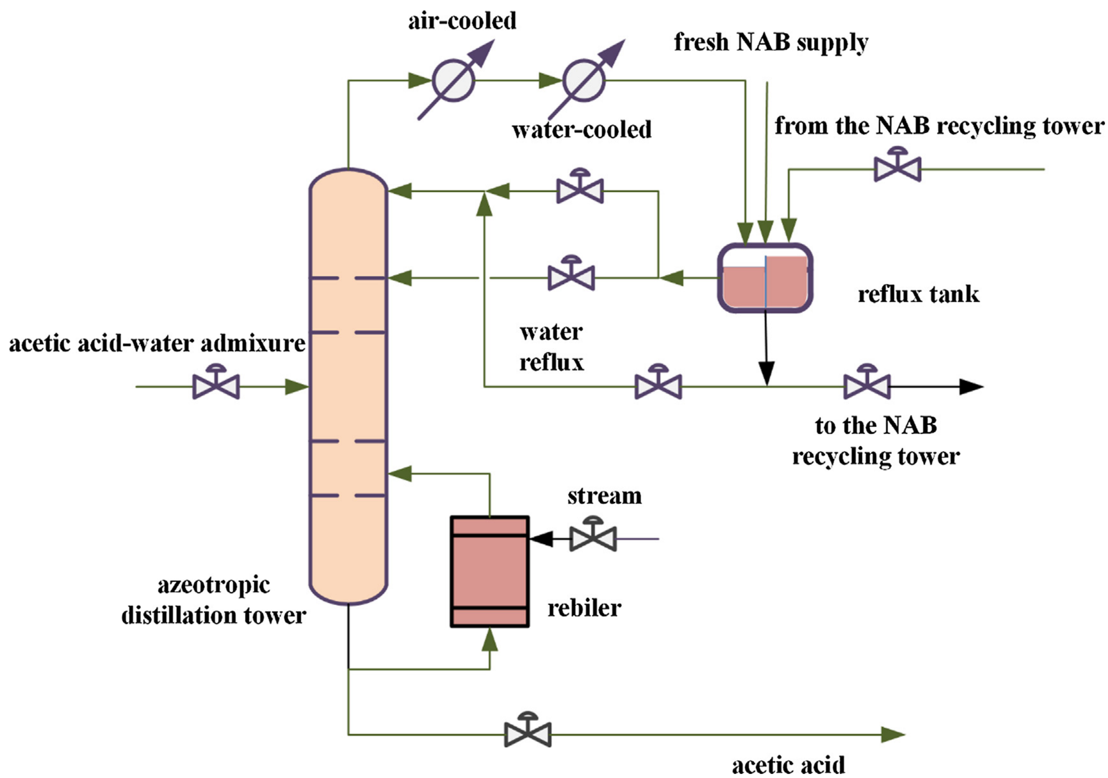

Fig. 2. A solvent dehydrating tower of the PTA production plant.

measure. Taking the conductivity of the top of the tower as an output variable, it can reflect the acetic acid content of the top of the tower.

There are 45 samples selected from the PTA production plant, and the input indicators and conductivity changes are shown in Fig. 3.

First, the input indicators are clustered by the AP clustering algorithm. The clustering results are shown in Fig. 4. The red circle represents the cluster center and the blue circle represents other input indicators. The temperature in the water reflux FC1502, FC1503, reboiler steam flow FC1507, TI1516, TI1517 and TC1501 in the tower are obtained, which have a great influence on the PTA production process.

These six indicators are used as inputs of the DEACM, and the conductivity of the tower top is used as the output of the DEACM to obtain improved efficiency values of the PTA production plant based on the AP-DEACM model. At the same time, the 17 indicators are used as the input of the traditional DEA model and the DEACM respectively, and the conductivity is taken as the output to obtain the efficiency value. The comparison of efficiency values of the general DEA model, the DEACM and the improved AP-DEACM in the PTA production plant is shown in Fig. 5.

It can be seen from Fig. 5 that the efficiency of 12 DMUs in the traditional DEA model is 1, including the sample 1, 5, 13, 14, 15, 16, 17, 23, 29, 37, 44 and 45. And the efficiencies of all DMUs are more than 0.96. When the efficiency values of DMUs are too many, there is no way to further distinguish between good and bad. In the DEACM, after the average of the cross-evaluation index and the self-evaluation index, the number of DMUs with the efficiency value of 1 is reduced,

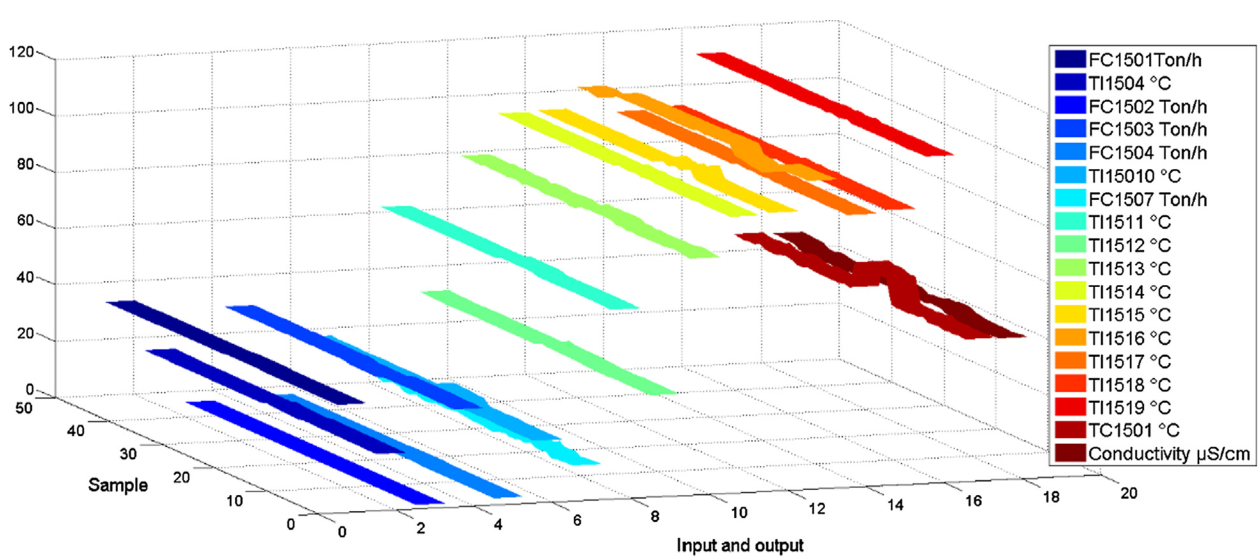

Fig. 3. Input and output of the PTA production plant.

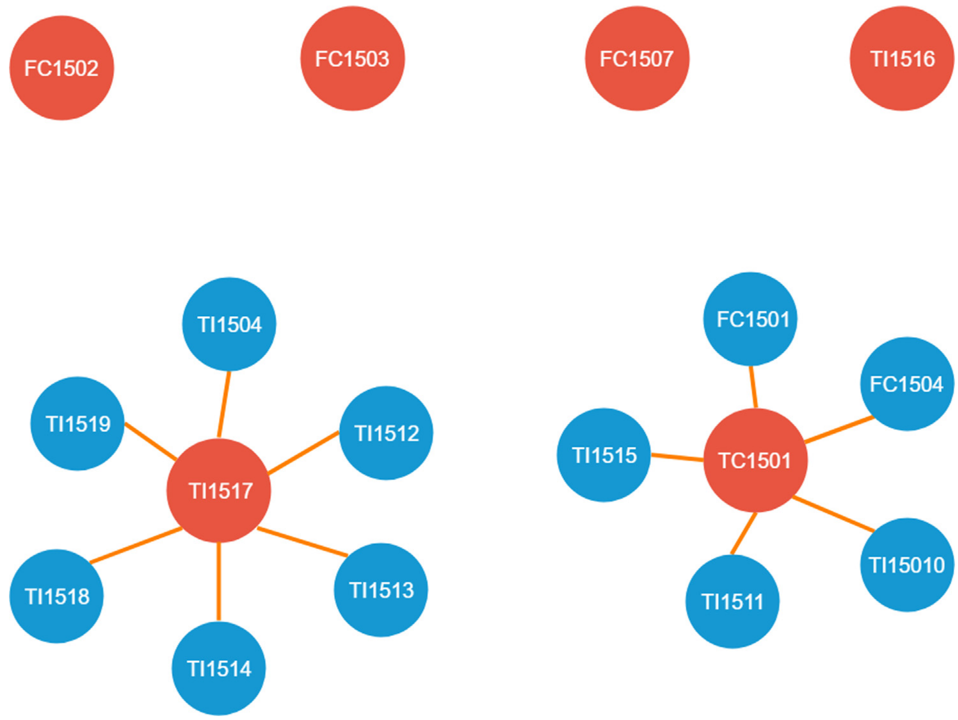

Fig. 4. The input indicators of the PTA plant based on the AP clustering algorithm.

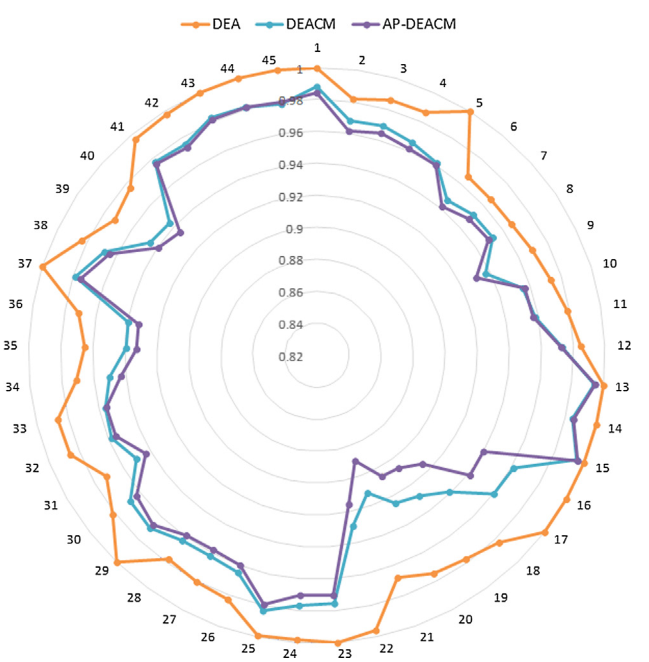

Fig. 5. The comparison result of the efficiency values in PTA production plants.

and the PTA production efficiency values are further differentiated. Among them, there are 26 DMUs with the efficiency value greater than 0.96, of which the first three samples are 15, 13, 1 and the last three samples are 19, 20 and 21. In the AP-DEACM, the efficiency value distribution is very similar to the DEACM. The efficiency value of 21 DMUs is greater than 0.96, and the samples of the first three are 15, 13 and 14 and the last three are 19, 20 and 21, which is basically consistent with those in the DEACM. However, the greater the standard deviation, the higher the discrimination of the DMU [43,44]. The efficiency value θ_i (i = 1, 2, …, 45) is taken as a variable. The standard deviation of the data is obtained based on Eq. (11)

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2} \tag{11}$$

where θ_i is the cross efficiency value for each sample, $\bar{\theta}$ is the average of the cross efficiency values of the samples, and n is the number of samples. The standard deviations of the traditional DEA model, the DEA cross model and the improved AP-DEACM are 0.0123, 0.0188 and 0.0237, respectively. Therefore, the AP-DEACM model is more accurate than the traditional DEA model, and the efficiency recognition is higher, which is more suitable for finding the improvement direction of the energy efficiency.

Based on the cross-evaluation efficiency values calculated by the proposed method, the sample 15 has the optimal production configuration and its data configuration can be used to improve input factors of other inefficiency samples. The comparison result of the input-output data of the ineffective sample 7, the sample 35 and the optimally production sample 15 in the PTA production plant is shown in Fig. 6.

According to the comparison result of the sample data in Fig. 6, the production parameters are adjusted for the sample 7 and the sample 35, and the result is shown in Fig. 7.

From Fig. 7, it can be concluded that for the sample 7, if the input index Water reflux FC1502 and Temperature inside the tower TC1501 are increased by 0.000088 Ton/h and 0.50116 °C, respectively, Water reflux FC1503, Reboiler steam flow FC1507, Temperature inside the tower TI1516 and Temperature inside the tower TI1517 are decrease by 0.30513 Ton/h, 0.19874 Ton/h, 2.4372 °C and 0.07196 °C, respectively, the output is increased by 2.099 μS/cm. For sample 35, the input indicators Water reflux FC1502 and Temperature inside the tower TI1517 are increased by 0.000159 Ton/h and 0.06306 °C, respectively, Water reflux FC1503, Reboiler steam flow FC1507, Temperature inside the tower TI1516 and Temperature inside the tower TC1501 are reduced by 0.0988 Ton/h, 0.89234 Ton/h, 2.4683 °C, 0.50863 °C, respectively, the output is increased by 2.1271 μS/cm. With the above improvements, Sample 7 and Sample 35 can achieve the efficient production. Other samples can also be efficiently produced according to the above analysis method. As shown in Fig. 3, considering the efficiency of all DMUs, the total change in conductivity caused by acetic

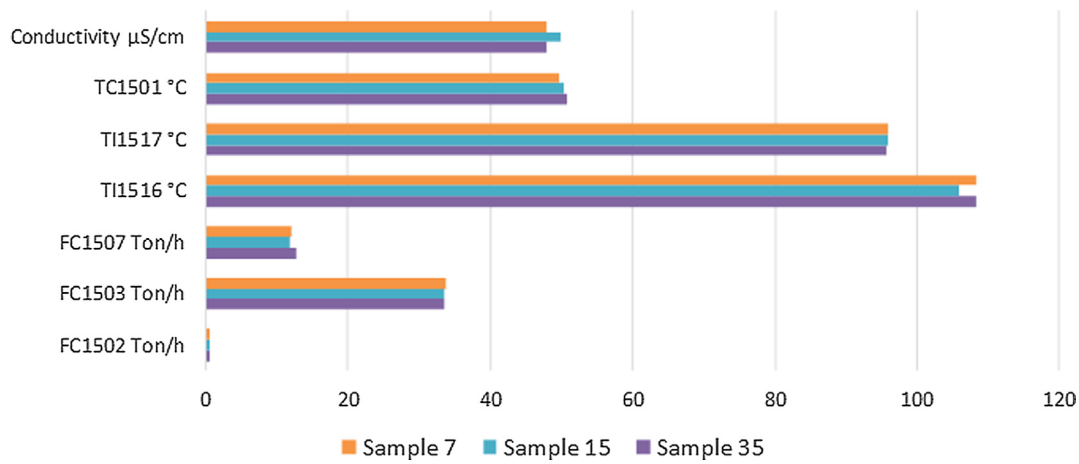

Fig. 6. The comparison result of the ineffective samples and the optimally production sample in the PTA production plant.

acid consumption is 60.937 μS/cm, and the average rate of change in conductivity is 2.78%. That is, the energy saving potential of PTA production plants is 2.78%.

4.2. Energy optimization and carbon emission reduction of the ethylene production system

The ethylene production process mainly includes two aspects of the cracking and the separation. The separation section includes: the rapid cooling, the compression and the separation [45,46]. Finally, high-additional chemical products such as the ethylene and the propylene are produced. A schematic flow diagram of an ethylene plant is shown in Fig. 8.

Overall, the production efficiency of ethylene production systems is closely related to crude oil, fuel, steam, electricity and water. The crude oil consists of naphtha, raffinate, hydrogoil, lhydr and C345, the fuel refers to gas, the steam includes the high press steam, the middle press steam and the low press steam, the water includes the circle water, the industry water and the boiler water, and the ethylene is the main output production.

There are 35 samples selected from the ethylene production plant, and the input indicators and the ethylene yield are shown in Fig. 9.

The input indicators are clustered by the AP clustering algorithm, and the clustering results are shown in Fig. 10. The red circle represents the cluster center and the blue circle represents other input indicators. Naphtha, hydrogoil, industry water electricity are obtained, which have a greater impact on the ethylene production.

These four factors are used as inputs of the DEACM, and the ethylene production is used as the output of the model, which is to obtain improved efficiency values of the ethylene production plant based on the AP-DEACM. Meanwhile, the 14 factors are used as the input of the traditional DEA model and the DEACM respectively, and the ethylene production is taken as the output, which is to obtain the efficiency value, respectively. The comparison result of efficiency values of the general DEA model, the DEACM and the AP-DEACM in the ethylene production plant is shown in Fig. 11.

It can be seen from Fig. 11 that the efficiency value of 32 DMUs in the traditional DEA model is 1 except sample 12, 18 and 21, which is more than 1/3 of the number of samples. When using the DEACM, the cross efficiency values of DMUs are greatly reduced and further differentiated. And the efficiency values of 15 DMUs are greater than 0.26. In the AP-DEACM, the result is very similar as the DEACM in terms of efficiency values. But the number of DMUs with the efficiency value greater than 0.58 is also 15, and the corresponding samples are identical, which indicate that the AP-DEACM is valid. According to Eq. (11), the standard deviations of the traditional DEA, the DEACM and the AP-DEACM are 0.0044, 0.0781 and 0.1503, respectively. And the AP-DEACM is more accurate than the traditional DEA model and the DEACM, and the efficiency value is more differentiated, which is more suitable for finding the direction of improving the ethylene yield.

Based on the cross-evaluation efficiency values calculated by the proposed method, the sample 28 has the optimal production configuration. Then it is used to guide the improvement input of other invalid samples. As shown in Fig. 12, the comparison result of the data of the

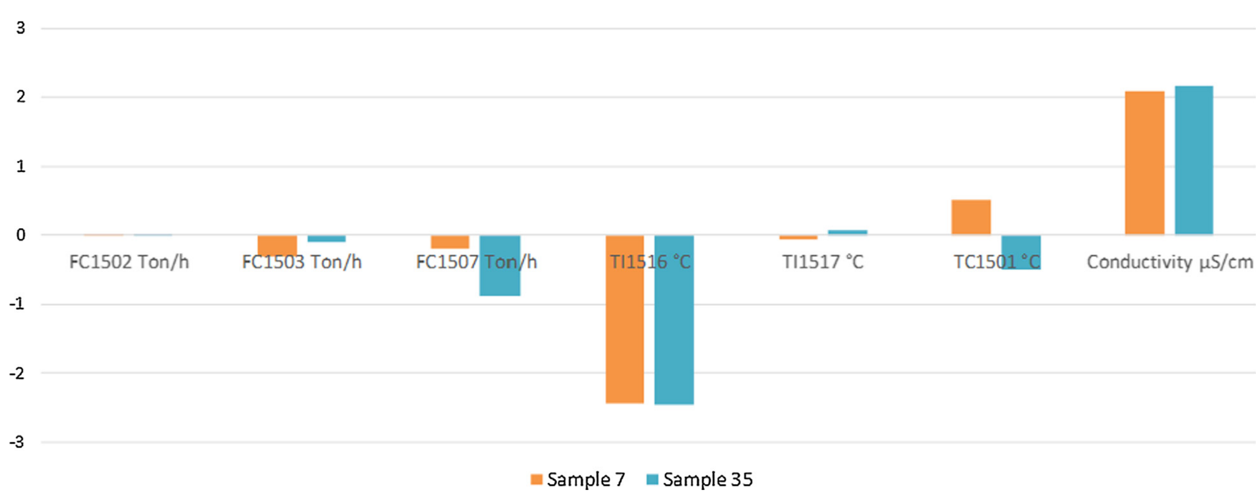

Fig. 7. Input factors and the conductivity adjustments for the sample 7 and the sample 35.

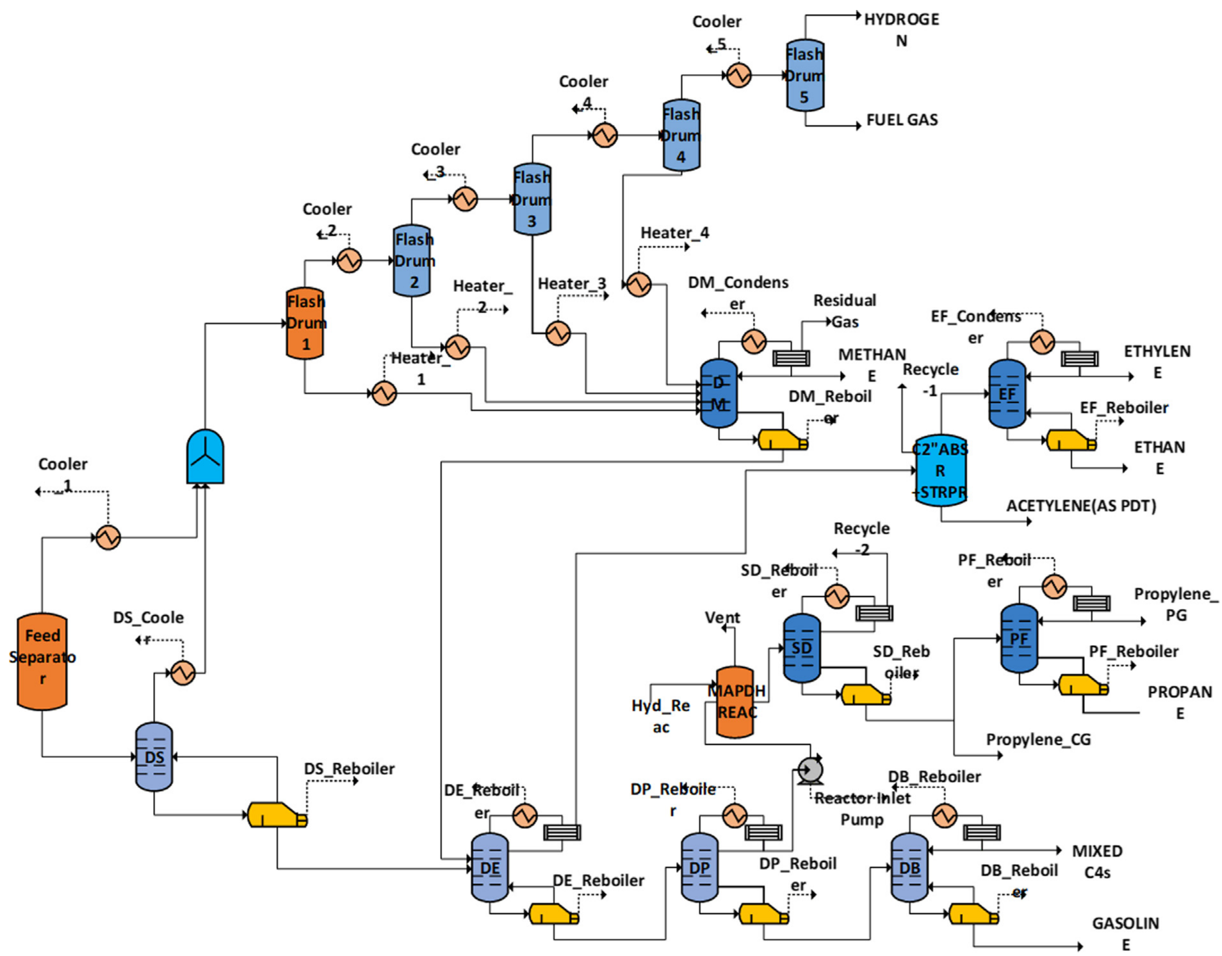

Fig. 8. The production process of a typical ethylene production plant.

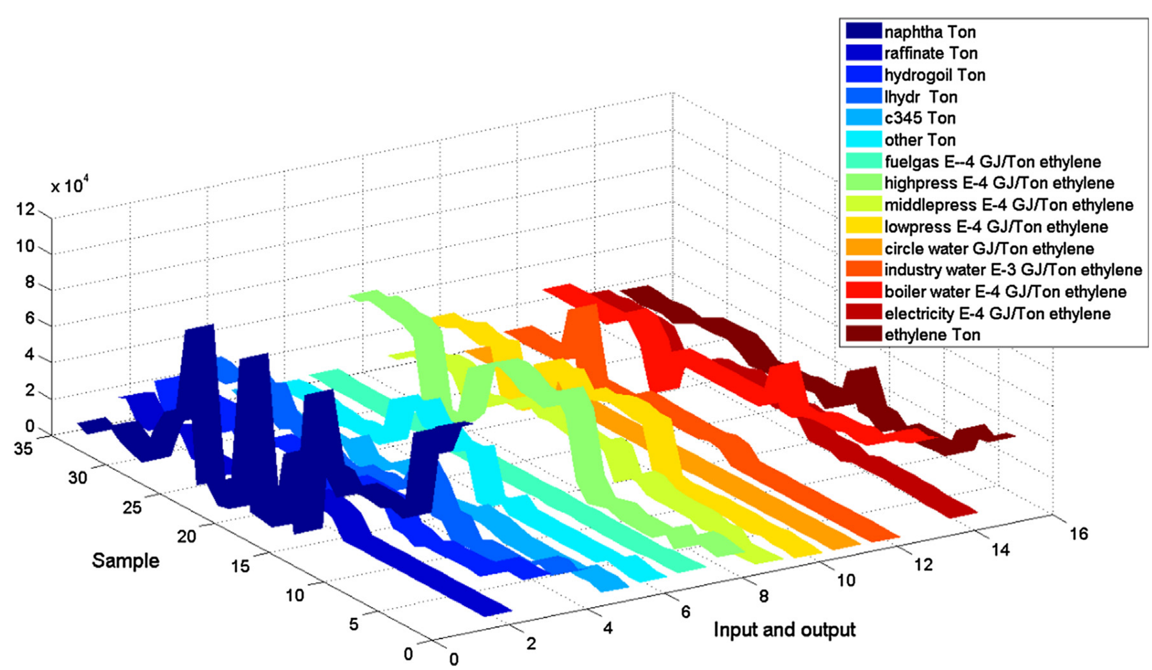

Fig. 9. Input and output of the ethylene production plant.

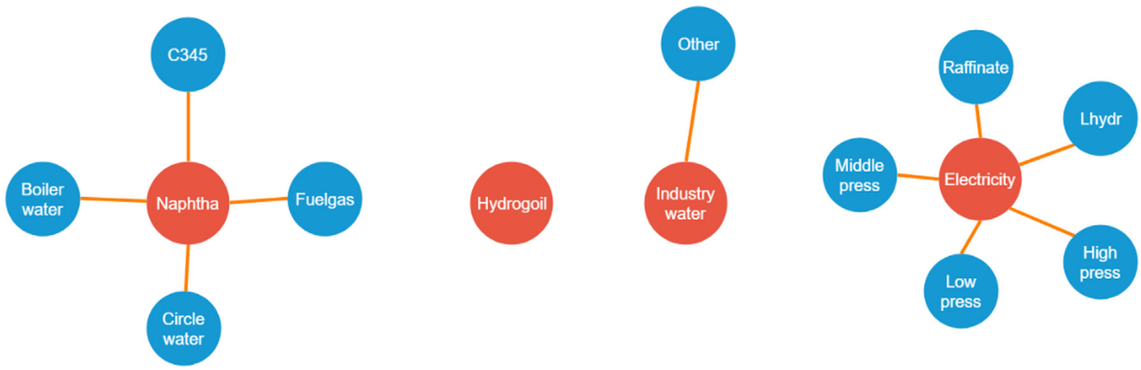

Fig. 10. The clustering results of ethylene production plants.

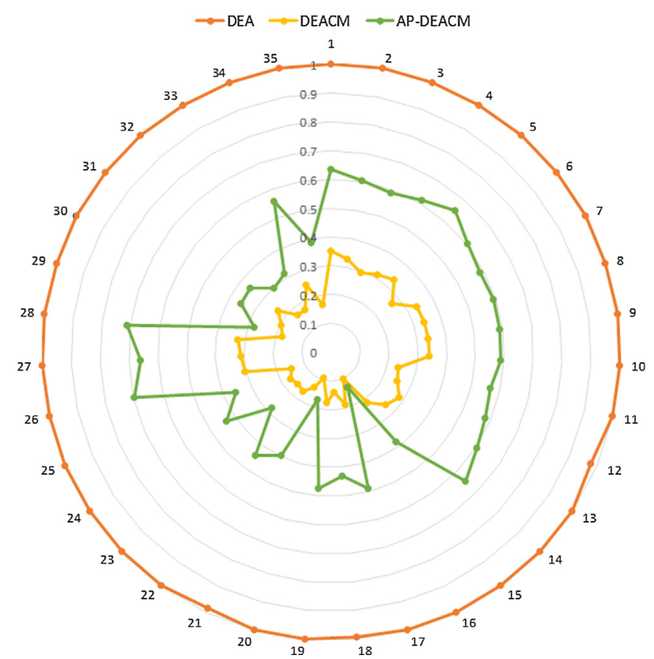

Fig. 11. Comparison results of the efficiency values in the ethylene production plant.

ineffective sample 20 and the optimally production sample 28 in the ethylene production plant is obtained.

According to the comparison result of the sample data in Fig. 12, the production input factors are adjusted for the sample 20 as shown in Fig. 13.

From Fig. 13, it can be concluded that for sample 20, if the input indicator naphtha, industry water, electricity are decreased by 65349.95 Ton, 5.6GJ/Ton ethylene and 0.93GJ/Ton ethylene respectively, hydrogoil of the crude oil is increased by 15,843 Ton, and the ethylene production will increase by 6022 Ton. With the above improvements, Sample 20 can achieve the efficient production. Other samples can also be achieved the maximum ethylene output based on the similar analysis. As shown in Fig. 9, the ethylene yield is about 733,843 Ton, but considering the increase in the efficiency of all DMUs, the ethylene yield can be increased by 9273 Ton, and the energy efficiency of the ethylene production plant is 1.26%.

Carbon dioxide emission is a very important indicator in the ethylene production process, so reducing carbon emission has become an important issue in the petrochemical industry. In actual production, the carbon emission coefficient can be estimated as shown in Eq. (12) [47].

$$C_i = \sum_{i=1}^{M} (x_{ij} \times c_j) \tag{12}$$

where C_i is the carbon dioxide emissions of DMU_i, x_{ij} is the input of energy j in sample i, c_j is the carbon emission coefficient of energy j, and M is the total number of energy input types. The carbon emission savings potential is estimated based on Eq. (13).

$$\omega = \frac{C_m}{W_m}$$

$$\varphi = \left(\frac{1}{M-1}\sum\nolimits_{i=1}^{M}\frac{C_i}{W_i} - \omega\right)/\omega i \neq m \tag{13}$$

where ω is the carbon emission of per ton ethylene produced by the optimal production sample, W_i is the ethylene yield of DMU_i, φ is the carbon emission savings potential, and m is the optimal sample code.

According to the general standard GB/T 2589-2008 [48] and the Development and Reform Office of Climate Change [2011] No. 1041 [49], the common energy carbon emission coefficients are shown in Table 1.

In ethylene production plants, carbon dioxide emissions are primarily related to the crude oil consumption. From Table 1, the carbon emission coefficient is 3.0202. Combined with Eqs. (12) and (13), the

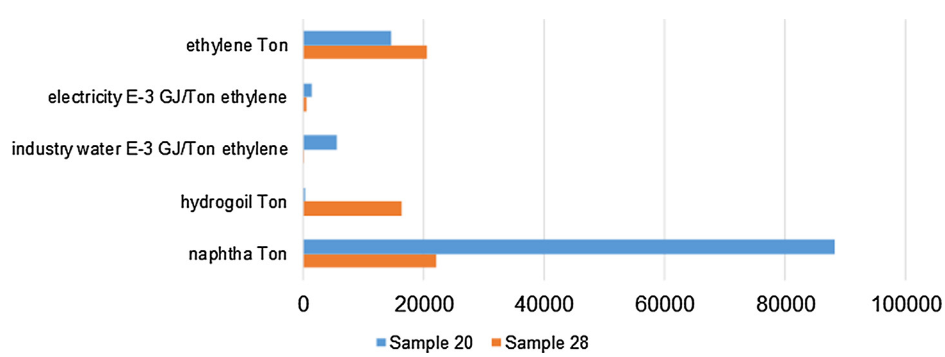

Fig. 12. The comparison result of the ineffective samples and the optimally production sample in the ethylene production plant.

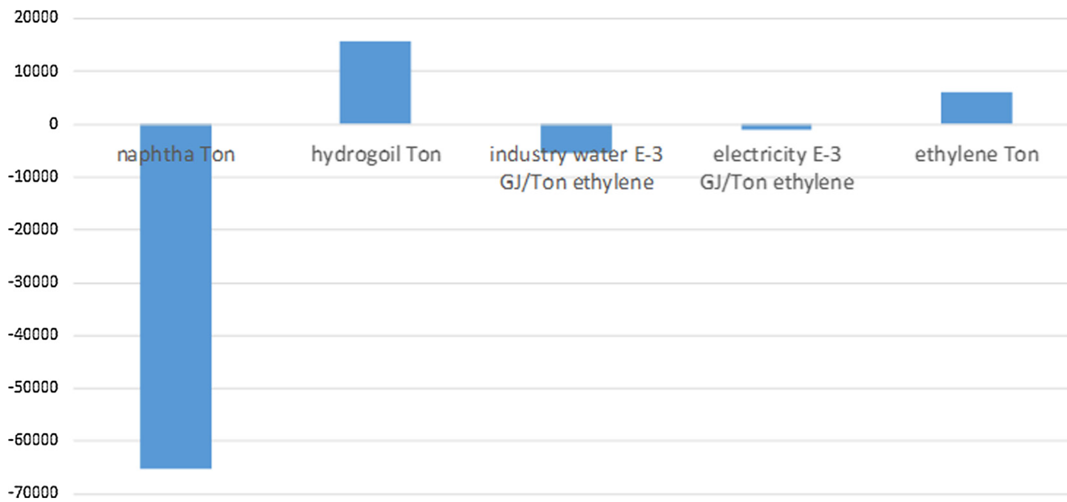

Fig. 13. Input factors and the ethylene yield adjustments for sample 20.

Table 1
Common carbon emission coefficients. (kg-CO_2/kg).

Energy	Carbon emission coefficient	Energy	Carbon emission coefficient
Raw coal	1.900 3	Kerosene	3.017 9
Coke	2.860 4	Diesel	3.095 9
Crude oil	3.020 2	Liquefied petroleum gas	3.101 3
Fuel oil	3.170 5	Refinery dry gas	3.011 9
Gasoline	2.925 1	Electricity	3.560 0

carbon emission savings potential is 3.62%.

5. Discussion

First, this paper proposes a novel AP-DEACM method. The original input data is clustered by the AP clustering algorithm to reduce the data dimensions and fuse the data with high similarity. Then the clustering results are used as inputs of the DEACM for evaluating the performance.

Second, the proposed method is applied to establish the performance evaluation and energy optimization model of the PTA production system and the ethylene production system in complex petrochemical industries. Compared with the general DEA and the DEACM, the results show that the AP-DEACM is less sensitive of the noise data than the DEA and the DEACM. Moreover, the energy saving potential of PTA and ethylene production plants is 2.78% and 1.26%, respectively. Considering the increased efficiency of all DMUs, the total change of conductivity in PTA production plant caused by acetic acid consumption is 60.937 μS/cm, and the ethylene production in ethylene production plants can be increased by 9273 Tons. Meanwhile, the average carbon savings potential of ethylene production plants is 3.62%.

Third, the proposed model also has some shortcomings. The clustering process needs to converge through continuous iterative calculations. When the amount of data is too large, the time complexity is high. Therefore, an appropriate algorithm, such as an adaptive artificial neural network (ANN), can be used to find the optimal damping coefficient λ, reducing the number of iterations, and achieving the purpose of reducing the time complexity.

6. Conclusion

This paper proposes a novel AP-DEACM method for evaluating the performance and optimizing the energy allocation in complex petrochemical industries. The AP-DEACM cannot only reduce the dimensionality of the high-dimensional data, but also filter out the redundant information. The clustering result of the AP is considered as an optimization index for processing optimization problems, and the energy state can be optimized and analyzed. The AP-DEACM method overcomes the effective phenomenon of excessive input indicators of the traditional DEA, and objectively evaluates the energy efficiency of PTA production systems and ethylene production systems in complex petrochemical industries. Finally, the performance evaluation and energy optimization model of PTA production systems and ethylene production systems are established by using the AP-DEACM. The results show that the proposed model can optimize the energy distribution and the product configuration to improve the energy efficiency of complex petrochemical industries. Moreover, the energy saving potential of PTA production plants and ethylene production plants are 2.78% and 1.26%, respectively, and the average carbon savings potential of ethylene production plants is 3.62%.

In our further work, we will integrate some deep learning methods to find the best damping coefficient and compare with the current work. In addition, there are more real-time data in the industry, so the processing of the data in every day or every hour, needs to be further researched. Moreover, the proposed model can be widely used in energy efficiency analysis and energy optimization in other complex petrochemical industries.

Conflict of interest

The authors declared that there is no conflict of interest.

Acknowledgement

This work is partly financial supported by the National Key Research and Development Program of China (2018YFB0803501), the National Natural Science Foundation of China (61603025 and 61533003) and the Fundamental Research Funds for the Central Universities (XK1802-4).

References

[1] Islam MT, Shahir SA, Uddin TMI, Saifullah AZA. Current energy scenario and future prospect of renewable energy in Bangladesh. Renew Sustain Energy Rev 2014;39(6):1074–88.
[2] Zhang M, Yu Y. Dehydration of ethanol to ethylene. Ind Eng Chem Res 2013;52(28):9505–14.
[3] Gao CY. Where is the road of China's ethylene industry competitiveness. Petrol Petrochem Today 2018;26(1):1–7.
[4] Geng ZQ, Qin L, Han YM, Zhu QX. Energy saving and prediction modeling of petrochemical industries: a novel ELM based on FAHP. Energy 2017;122:350–62.
[5] Li Z, Zhong WM, Wang XQ, Luo N, Qian F. Control structure design of an industrial crude terephthalic acid hydropurification process with catalyst deactivation.

Comput Chem Eng 2016;88(6):1–12.
[6] Li J. Research on the status and development trend of China's PTA industry. Strait Sci Technol Ind 2017;3:144–5.
[7] Bailey JA, Gordon R, Burton D, Yiridoe EK. Energy conservation on Nova Scotia farms: baseline energy data. Energy 2008;33(7):1144–54.
[8] Geng ZQ, Gao HC, Wang YQ, Han YM, Zhu QX. Energy saving analysis and management modeling based on index decomposition analysis integrated energy saving potential method: application to complex chemical processes. Energy Convers Manage 2017;145:41–52.
[9] Fetanat A, Shafipour G. A hybrid method of LMDI, symmetrical components, and SFA to estimate the distribution of energy-saving potential with considering of unbalanced components in decomposition analysis. Energy Efficiency 2017;10(4):1041–59.
[10] Nassiri SM, Singh S. Study on energy use efficiency for paddy crop using data envelopment analysis (DEA) technique. Appl Energy 2009;86(7):1320–5.
[11] Afzal MNI, Lawrey R. A measurement framework for knowledge-based economy (KBE) efficiency in ASEAN: a data envelopment (DEA) window approach. Int J Bus Manage 2012;7(18):57–68.
[12] Hernándezsancho F, Molinossenante M, Salagarrido R. Energy efficiency in Spanish wastewater treatment plants: a non-radial DEA approach. Sci Total Environ 2011;409(14):2693–9.
[13] Song ML, Zhang LL, Liu W, Fisher R. Bootstrap-DEA analysis of BRICS' energy efficiency based on small sample data. Appl Energy 2013;112:1049–55.
[14] Bi GB, Shao YY, Song W, Yang F, Luo Y. A performance evaluation of China's coal-fired power generation with pollutant mitigation options. J Cleaner Prod 2018;171:867–76.
[15] Azadeh A, Rahimi-Golkhandan A, Moghaddam M. Location optimization of wind power generation–transmission systems under uncertainty using hierarchical fuzzy DEA: a case study. Renew Sustain Energy Rev 2014;30(2):877–85.
[16] Wen M, You C, Kang R. A new ranking method to fuzzy data envelopment analysis. Comput Math Appl 2010;59(11):3398–404.
[17] Song CX, Li MJ, Zhang F, He YL, Tao WQ. A data envelopment analysis for energy efficiency of coal-fired power units in China. Energy Convers Manage 2015;102:121–30.
[18] Hua J, Ren J, Xu M. Research on performance evaluation of regional CO2 emission reduction in China based on three-stage DEA. Resour Sci 2013;35(7):1447–54.
[19] Han YM, Geng ZQ, Gu XB, Zhu QX. Energy efficiency analysis based on DEA integrated ISM. Eng Appl Artif Intell 2015;45:80–9.
[20] Geng ZQ, Wang ZK, Zhu QX, Han YM. Energy efficiency evaluation method based on IDA-DEA and its applications in ethylene industries. CIESC J 2017;68(3):910–5.
[21] Han YM, Long C, Geng ZQ, Zhang KY. Carbon emission analysis and evaluation of industrial departments in China: an improved environmental DEA cross model based on information entropy. J Environ Manage 2017;205:298–307.
[22] Sexton TR, Silkman RH, Hogan AJ. Data envelopment analysis: critique and extensions. New Dir Eval 2010;1986(32):73–105.
[23] Chen W, Zhou K, Yang S. Evaluation of China's electric energy efficiency under environmental constraints: a DEA cross efficiency model based on game relationship. J Cleaner Prod 2017;164:38–44.
[24] Cook WD, Zhu J. DEA Cobb-Douglas frontier and cross-efficiency. J Oper Res Soc 2014;65(2):265–8.
[25] Gao QM, Zhang QY. Comprehensive evaluation and decision-making for urban power system planning using a cross efficiency model based on SE-DEA. Power Syst Prot Control 2011;39(8):60–4.
[26] 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;83:685–95.
[27] Han YM, Geng ZQ, Gu XB, Wang Z. Performance analysis of China ethylene plants by measuring Malmquist production efficiency based on improved DEA cross model. Ind Eng Chem Res 2015;54(1):272–84.
[28] An QX, Meng FY, Xiong BB. Interval cross efficiency for fully ranking decision making units using DEA/AHP approach. Ann Oper Res 2018;271(2):297–317.
[29] Wu J, Liang L. A multiple criteria ranking method based on game cross-evaluation approach. Ann Oper Res 2012;197(1):191–200.
[30] Azadeh A, Amalnick MS, Ghaderi SF, Asadzadeh SM. An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. Energy Policy 2007;35(7):3792–806.
[31] Gong SX, Shao C, Zhu L. Energy efficiency evaluation in ethylene production process with respect to operation classification. Energy 2016;118:1370–9.
[32] Rotem Y, Wachs A, Lewin DR. Ethylene compressor monitoring using model-based PCA. AIChE J 2010;46(9):1825–36.
[33] Wang F, Li K, Duić N, Mi Z, Hodge BM, Shafie-khah M, et al. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. Energy Convers Manage 2018;171:839–54.
[34] Hu J, Wu M, Chen X, Du S, Zhang P, Cao WH, et al. A multilevel prediction model of carbon efficiency based on differential evolution algorithm for iron ore sintering process. IEEE Trans Ind Electron 2018;65(11):8778–87.
[35] Wei ZX, Wang YX, He SL, Bao JD. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. Knowl-Based Syst 2016;116:1–12.
[36] Gao JL, Yao F, Xu YP, Sun GS, Zheng CZ, Qi SH, et al. PMA partition method of water distribution network combined with Gragh theory. Procedia Eng 2017;186:278–85.
[37] Zagouras A, Kolovos A, Coimbra CFM. Objective framework for optimal distribution of solar irradiance monitoring networks. Renew Energy 2015;80:153–65.
[38] Chen B, Song K. Multi-model modeling of O2 content in flue gas of power plant based on affinity propagation clustering. Appl Mech Mater 2013;448–453:3344–9.
[39] Dueck D, Frey BJ. Non-metric affinity propagation for unsupervised image categorization. IEEE international conference on computer vision. 2007. p. 1–8.
[40] Foster B, Bagci U, Xu ZY, Dey B, Luna BM, Bishai WR, et al. Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models. IEEE Trans Biomed Eng 2014;61(3):711–24.
[41] 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 propagations. Energy Convers Manage 2019;180:240–9.
[42] Chen ZS, Zhu B, He YL, Yu LA. A PSO based virtual sample generation method for small sample sets: applications to regression datasets. Eng Appl Artif Intell 2017;59:236–43.
[43] Han DQ, Han CZ, Deng Y, Yang Y. Weighted combination of conflicting evidence based on evidence variance. Acta Electronica Sinica 2011;39(3A):153–7.
[44] Yousefi S, Soltani R, Saen RF, Pishvaeed MS. A robust fuzzy possibilistic programming for a new network GP-DEA model to evaluate sustainable supply chains. J Cleaner Prod 2017;166:537–49.
[45] Schulz EP, Bandoni JA, Diaz MS. Optimal shutdown policy for maintenance of cracking furnaces in ethylene plants. Ind Eng Chem Res 2006;45(8):2748–57.
[46] Tarafder A, Lee BCS, Ray AK, Rangaiah GP. Multiobjective optimization of an industrial ethylene reactor using a nondominated sorting genetic algorithm. Ind Eng Chem Res 2005;44(1):124–41.
[47] Ye B, Jiang JJ, Li CS, Miao LX, Tang J. Quantification and driving force analysis of provinciallevel carbon emissions in China. Appl Energy 2017;198:223–38.
[48] China Standards: the general computing guide of special energy consumption (GB/T2589-2008); 2008.
[49] China Standards: guidelines for the preparation of provincial greenhouse gas inventories (Development and Reform Office Climate [2011] No. 1041); 2011.