Journal of Analytical and Applied Pyrolysis 125 (2017) 347–355

Contents lists available at ScienceDirect

Journal of Analytical and Applied Pyrolysis

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

Linear optimization fusion model based on fuzzy C-means: Case study of energy efficiency evaluation in ethylene product plants

Yongming Han[a,b], Zhiqiang Geng[a,b,*], Yixin Qu[c], Qunxiong Zhu[a,b,*]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education in China, Beijing 100029, China
[c] High-Tech Research Institute, Beijing University of Chemical Technology, Beijing 100029, China

ARTICLE INFO

Article history:
Received 24 December 2016
Received in revised form 16 February 2017
Accepted 5 March 2017
Available online 14 March 2017

Keywords:
Ethylene production plant
Energy efficiency analysis
Linear optimization fusion
Fuzzy C-means clustering

ABSTRACT

Ethylene production shares the common characteristics of different techniques and scales, high dimension and multi-correlation. However, the existing methods do not take the influence of the technology, scale and materials into account. In order to analyze the energy efficiency of ethylene plants better, this paper proposes a linear optimization fusion model based on fuzzy C-means (FCM) clustering. The FCM clustering is used to obtain the corresponding classification and clustering centers of the large scale time series data of the ethylene production process. Then the mean squares of time-series variables are calculated by various clustering centers. The multidimensional energy efficiency benchmarking of ethylene plants can be extracted by the fusion weighting vectors in linear optimization fusion model. The experimental results indicate that the proposed method intuitively reveals the running state and the operating level of ethylene plants and accurately identify the reason of energy consumption for improving energy efficiency.

1. Introduction

Energy is one of the most important parts of modern economy and the cornerstone of modern civilization. With the rapid development of economy, energy problems have become increasingly prominent. The energy conservation and emissions reduction has become one of the common themes of the world development today, especially in the rapid developing ethylene industries. Ethylene production level often serves as one of the main judgments for the industrial development level of a country. According to statistics, the average fuel plus power consumption (standard oil) was 571.39 kg/t for producing a ton of ethylene when China Petrochemical Corporation's ethylene production was 10420 kt/a in 2014 [1]. And China National Petroleum Corporation's ethylene production was 4976 kt/a and the standard oil was 616.7 kg per ton of ethylene produced in 2014 [2]. Meanwhile, More than 50% of the ethylene plants operating costs are derived from the cost of energy consumption of ethylene [3]. The statistics indicated that comparing with the developed countries, the ethylene energy consumption of China is significantly higher. Therefore, there is enough room to improve energy efficiency of producing the ethylene in China. Moreover, the study of energy efficiency evaluation of ethylene plants will achieve the environment and the sustainable development of the petrochemical industry. In order to adopt energy-saving measures effectively, it is an effective way to assess the energy efficiency of the existing ethylene plants objectively and established ethylene energy efficiency evaluation indices effectively.

At present, Enterprises usually utilize the average strategy method and the optimal index method to establish the benchmark efficiency analysis [4]. However, the energy saving knowledge did not been taken into account by these methods, so they cannot give a performance efficiency value benchmarking of optimal factors and indices to guide the improvement of the real-world energy efficiency status. Additionally, researchers reduced the energy consumption of the ethylene production industry by advanced process integration and their operation optimization [5,6]. Yousefi et al. studied energy saving by optimizing the design problem of real-world compact heat exchangers with variable heat duties [7]. Fu and Xu studied on energy consumption and emission generation for an ethylene plant under different start-up strategies by plant-wide dynamic simulations [8]. However, the above methods did not take the economic cost of reforming industry plants into consideration. When the special energy consummation (SEC) per ton ethylene is used to evaluate energy efficiency, the energy consump-

* Corresponding authors at: Beijing University of Chemical Technology, 15# Beisanhuan East Rd., Chaoyang District, Beijing 100029, China.
E-mail addresses: gengzhiqiang@mail.buct.edu.cn (Z. Geng), zhuqx@mail.buct.edu.cn (Q. Zhu).

http://dx.doi.org/10.1016/j.jaap.2017.03.007
0165-2370/

Nomenclature

FCM	fuzzy C-means
SEC	Special Energy consummation
AHP	analytic hierarchy process
DEA	Data Envelopment Analysis
c	the number of clusters
m	the fuzzy weighted index
x_k	The value of the k-th sample
R^s	the dimensions of the vectors
u_{ik}	the membership degree of the k-th sample belonging the i-th clustering center
v_i	s dimension vector
d_{ik}	the Euclidean distance between the k-th fuzzy group and the i-th clustering center
ε	the iteration stop threshold
t	the iteration number
X	The time-series data collection
x_{ij}	The i row j column value of X
$x_j^{\max}$	$\max\left\{x_{1j}, x_{2j}, \ldots, x_{tj}\right\}$
$x_j^{\min}$	$\min\left\{x_{1j}, x_{2j}, \ldots, x_{tj}\right\}$
N_i	the length of the time-series data contained in i-th clustering center
W^T	the comprehensive weighted vector
A	the column vector whose elements are all 1
σ_{ij}	the mean square error of the i-th class j-th parameter
R_{kk}	the n order symmetric matrix
W_{opt}	the linear weight vector

tion of ethylene plants is susceptible to be affected by feedstock composition, technological innovation, scale transformation, load changes and so on. Thus, in order to obtain SEC and parameter values of the above factors at the same time, it is necessary to use the actual operation time-series data of ethylene plants.

The time-series data of ethylene production plants share the characteristics of high dimension, multi-correlation, and containing noise [9,10], so the focus of time-series data analysis is to use the appropriate algorithm to find their similarities, reduce dimension reasonably, remove noise data simultaneously and decrease data correlation. The Euclidean distance is used in the vast majority of the typical similarity measures [9], and it is an effective sequence analysis method for time series to determine their physical shape and developing trend by cluster analysis methods adopting the Euclidean distance as the similarity measure. Data fusion is also an effective way to reduce the dimension. Zhou et al. took linear minimum variance estimation as the optimal criteria of fusion estimation and proposed data optimum linear data fusion to satisfy two addition principles [11]. According to the theorem, it is not difficult to find that researching on linear minimum variance optimization fusion for large time-series data can extract effectively and embed synthetically the information and knowledge related to the time-series data (such as extracting benchmark information contained in the time-series data, etc.). With the data amount increasing, the results of the linear minimum variance optimization fusion become reliable and accurate. However, when K-means clustering linear minimum variance clustering algorithm is used for energy efficiency analysis, it is necessary to consider choosing K-means initial value. While the obtained weight values are negative, deviations of fusion results are larger [12]. The data fusion method based on association rules achieved good application results in the ethylene plant energy efficiency analysis [13]. But the method did not take the role of crude factor on energy consumption indicators into consideration. Fuzzy Analytic Hierarchy Process introduced expert experiences allocation weights in the evaluation of large-scale equipment (such as ethylene plants) state, resulting in lack of objectivity [14]. Moreover, the method based on the Analytic Hierarchy Process (AHP) and Data Envelopment Analysis (DEA) model has been widely applied to efficiency evaluation of agriculture, logistics and other industries [15–17]. However, the crude oil was not taken into consideration in ethylene performance [18,19].

To avoid the problem of the negative fusion weights, the paper proposes a linear optimization fusion model based on FCM. The optimal linear data fusion addition principle and the linear variance of the time-series data as the optimal criterion are applied. First, each time-series variable variance can be calculated based on various types of C-means clustering centers. Second, combined with the detailed derivation and proof of multivariate time-series data linear variance optimization fusion algorithm, the linear variance optimization fusion weighted vectors is used to fuse all cluster centers. Finally, the energy efficiency value features contained in the time-series multi-dimension data are extracted. Then the proposed method is applied to evaluate the energy efficiency of ethylene plants. The results indicate that this method intuitively reveals the running state and the operating level of various ethylene plants under different techniques and scales, and accurately identify the main reason of energy consumption to specify the main direction to improve energy efficiency.

Section 1 introduces the research status of energy efficiency in ethylene plants. Section 2 provides the details of fuzzy C-means clustering analysis and mean variance. The multivariate data optimization fusion based on fuzzy C-means is described in Section 3. Section 4 presents the application example about the energy efficiency value of ethylene plants. Finally, the conclusions are given in Section 5.

2. Calculation of fuzzy C-means clustering analysis and mean variance

In practical applications, the amount of the time-series data requiring fusion are generally large (probably GB bytes, even TB bytes), and there is the case of time-series data related to several variables. So it is necessary to make a clustering analysis of time-series data, effectively extracting the relevant information embedded in the time-series data. For most of time-series data, it is not always feasible to segment simultaneously on the time and physical quantity size (such as energy consumption, etc.), sometimes even not necessary. As time changes, there will be some changes and abnormal conditions for the variables in the state. The clustering analysis method is applied to analyze the sequence and heterogeneous data [20], and can effectively eliminate the scatter points of time-series data. Therefore, the FCM clustering algorithm is adopted to preprocess the time-series data.

Traditional hard clustering algorithms, such as K-means algorithm, are sensitive to the initial clustering centers [21,22], and clustering results fluctuate with different initial inputs. However, the FCM clustering method is a soft clustering method which uses the membership degree to determine to the variables belonging to a clustering degree [23]. The main idea of the FCM clustering algorithm is to fuzz the definition of classic division. The FCM has two parameters: one is the number of clusters c; the other is fuzzy weighted index m. c> 1, and the c value is usually much smaller than the total number of clustering samples. The m is also known as the smoothing parameter. If the m value is too large, the clustering effect is not good. If the m value is too small, the clustering effect is very close to the hard clustering algorithm. The m usually takes [1.5, 2.5] [24].

FCM clustering algorithm divides n vectors $x_k \subset R^s$ into c groups R_{kk}, s indicating the dimension of the vector x_k, and the clustering center of each group is obtained. The data are classified by solving the membership matrix $U = \{u_{ik}\}_{c*n}$, where u_{ik} is the membership degree of the k-th sample belonging the i-th clustering center, its value in the interval [0,1], and the sum of the membership degree of each sample is 1. The specific mathematical expression is as follows. Set

$$V = [v_1, v_2, \ldots, v_c]^T \tag{1}$$

as c clustering centers, where $v_i = [v_{i1}, v_{i2}, \ldots, v_{is}]$ is s dimension vector ($i = 1, 2, \ldots, c$), and then the generalized form of the objective function m is

$$J(U, v_1, \ldots, v_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 \tag{2}$$

where $d_{ik} = \sqrt{\sum_{q=1}^{s}\left(x_{kq} - v_{iq}\right)^2}$ is the Euclidean distance between the k-th fuzzy group and the i-th clustering center, the fuzzy weighted index $m \in (1, +\infty)$, $2 \le c \le n$, and $\sum_{i=1}^{c} u_{ik} = 1$, $u_{ik} \in (0, 1)$.

The FCM algorithm is carried out iteratively by optimizing J (U, V), and the specific steps are as follows:

(1) Set c and m, and set the iteration stop threshold $\varepsilon > 0$ and the iteration number $t = 0$, respectively. And the clustering center V is initialized.
(2) The membership matrix is updated by Eq. (2).
(3) The clustering center is updated by Eq. (1).
(4) If $V^{(k+1)} - V^{(k)} < \varepsilon$, then the algorithm stops, otherwise repeat step (2) and (3), and set $k = k + 1$.

The time-series data collection X of m variables is expressed as Eq. (3), where t is the length of sequence.

$$X = \begin{array}{c c} & \begin{array}{cccc} x_1 & x_2 & \ldots & x_m \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ \ldots \\ X_t \end{array} & \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ x_{t1} & x_{t2} & \ldots & x_{tm} \end{bmatrix} \end{array} \tag{3}$$

Generally, each variable dimension of time-series data is not the same, so the values are not comparable between the variables. It is necessary to normalize the time-series dataset before clustering. The common conversion method is proportion conversion. When using this method, it needs to pay attention to the theme described by the time-series data. For this theme, the effect types of different variables may be different. Some variables have the positive effect on the theme [12], using the following conversion formula:

$$x'_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{4}$$

Some variables have the negative effect on the theme, and then use the following conversion formula:

$$x'_{ij} = 1 - \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}} \tag{5}$$

Where, $x_j^{\max} = \max\{x_{1j}, x_{2j}, \ldots, x_{tj}\}$, $x_j^{\min} = \min\{x_{1j}, x_{2j}, \ldots, x_{tj}\}$, $i = 1, 2, \ldots, t$, $j = 1, 2, \ldots, m$, where the data zero uses a very small positive integer instead, such as 0.000001.

The above conversions can ensure the information matrix used in fuzzy C-means clustering is a positive definite matrix, in order to ensure the accuracy and understandability of clustering results.

Adopting the above algorithms, the results of clustering centers of n classes can be obtained by applying the fuzzy C-means clustering algorithm to normalized multivariate time-series data, as the following in Eq. (6).

$$C = \begin{array}{c c} & \begin{array}{cccc} x_1 & x_2 & \ldots & x_s \end{array} \\ \begin{array}{c} C_1 \\ C_2 \\ \ldots \\ C_n \end{array} & \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{is} \\ c_{21} & c_{22} & \ldots & c_{2s} \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & c_{n2} & \ldots & c_{ns} \end{bmatrix} \end{array} \tag{6}$$

With reference to calculation formula of mean square error, by using the clustering centers of n classes in Eq. (6) and all time-series data contained in various classes, we can obtain mean square error calculation formula based on the clustering center, as shown in Eq. (7), to calculate the mean square error of each variable for each clustering center.

$$\sigma_{ij} = \sqrt{\frac{1}{N_i}\sum_{l=1}^{N_i}\left(x'_{ij} - \overline{x'_{ij}}\right)^2} = \sqrt{\frac{1}{N_i}\sum_{l=1}^{N_i}(x'_{ij} - c_{ij})^2} \tag{7}$$

where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$, and N_i is the length of the time-series data contained in i-th clustering center.

3. Multivariate data optimization fusion based on fuzzy C-means

The fusion problem of the fuzzy C-means clustering center, shown in Eq. (6), is actually a multivariate fusion which describes the same theme but different time, so it is not necessary to use different weights to dispose the different variables by the optimal state estimation. The each clustering center is set as with a “sensor”, a comprehensive weighted vector $W^T = \begin{bmatrix} w_1 & w_2 & \ldots & w_n \end{bmatrix}$ is chosen for each clustering center weighted fusion, then the linear optimization fusion problem of the matrix C can be expressed as: $Y = W^T C$, where the multivariate vector of the fusion result.

$$Y = \begin{bmatrix} y_1 & y_2 & \ldots & y_m \end{bmatrix} \tag{8}$$

To solve the minimal value of matrix quadratic optimization problem: $W_{opt} = arg\min[P]$, where

$$P = E\left\{W^T[C - E(C)]\right\}^2 = E\left\{W^T[C - E(C)][C - E(C)]^T W\right\} \tag{9}$$

under the conditions of convex linear constraint equation

$$W^T A = 1 \tag{10}$$

where A is a column vector whose elements are all 1.

We can obtain P by Eq. (9).

$$P = W^T R_{kk} W \tag{11}$$

where

$$R_{kk} = E\left\{[C-E(C)][C-E(C)]^T\right\}$$

$$= \begin{pmatrix} c_{11}-E(c_{11}) & \dots & c_{1m}-E(c_{1m}) \\ \dots & \dots & \dots \\ c_{n1}-E(c_{n1}) & \dots & c_{nm}-E(c_{nm}) \end{pmatrix} \begin{pmatrix} c_{11}-E(c_{11}) & \dots & c_{1m}-E(c_{1m}) \\ \dots & \dots & \dots \\ c_{n1}-E(c_{n1}) & \dots & c_{nm}-E(c_{nm}) \end{pmatrix}^T$$

$$= \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1m} \\ \dots & \dots & \dots \\ \sigma_{n1} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1m} \\ \dots & \dots & \dots \\ \sigma_{n1} & \dots & a_{nm} \end{pmatrix}^T = \begin{bmatrix} \sum_{q=1}^{m}\sigma_{1q}\sigma_{1q} & \sum_{q=1}^{m}\sigma_{1q}\sigma_{2q} & \dots & \sum_{q=1}^{m}\sigma_{1q}\sigma_{nq} \\ \sum_{q=1}^{m}\sigma_{2q}\sigma_{1q} & \sum_{q=1}^{m}\sigma_{2q}\sigma_{2q} & \dots & \sum_{q=1}^{m}\sigma_{2q}\sigma_{nq} \\ \dots & \dots & \dots & \dots \\ \sum_{q=1}^{m}\sigma_{nq}\sigma_{1q} & \sum_{q=1}^{m}\sigma_{nq}\sigma_{2q} & \dots & \sum_{q=1}^{m}\sigma_{nq}\sigma_{nq} \end{bmatrix} \quad (12)$$

$$== \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}$$

where σ_{ij} is the mean square error of the i-th class j-th parameter, obtained by Eq. (7).

For the n order symmetric matrix R_{kk}, according to the product square root method (geometric mean method) [18,25], the feature vector $W = (w_1, w_2, \dots, w_n)^T$ can be obtained. The specific process is as follows:

$$w_i = r_i/r \ (i = 1, 2, \dots, n) \quad (13)$$

where $r_i = (\prod_{j=1}^{n} r_{ij})^{\frac{1}{n}} \quad (i = 1, 2, \dots, n), r = \sum_{i=1}^{n} r_i (i = 1, 2, \dots, n)$

Using W to fusing the program, we get the fusion data of plant energy efficiency value:

$$Y = W_{opt}^T C \quad (14)$$

Diagrammatically, Fig. 1 shows the whole process of linear optimization fusion model based on fuzzy C-means.

So far it is not difficult to find the linear weight vector W_{opt} can obtain the fusion results, and avoid the problem in the literature [12] that the obtained minimum variance based on K-means is negative fusion weights. Although the literature [12] proposed two methods to process the negative weights obtained by the analysis results, it changed the effect of the minimum variance fusion results. Based on the product square root method, this paper eliminates the effect of negative weights, and gets good fusion results. Meanwhile, the other types of data (good data) can obtain greater weights by Eq. (13), so that the fusion results are more reasonable.

4. Case study: energy efficiency evaluation application in ethylene plants

A typical framework of the typical ethylene plant flow sheet is shown in Fig. 2 [8]. Two types of feedstock are used in the typical ethylene plant for thermal cracking: light naphtha and heavy naphtha. Two for cracking heavy feed and the other five for cracking light feed are available. The cracked gas from furnaces is sent to oil quench and water quench towers sequentially. In the quench system, heat is removed by circulating quench oil and quench water while cracked gas is cooled and partially condensed. The quench tower overhead vapor is compressed in a five-stage cracked gas compression (CGC) section. Between the 3rd and 4th compression stages, the cracked gas is treated by caustic water/wash to remove acid gases generated in the cracking heaters. The cracked gas from the 5th stage compressor is dried firstly, and then chilled against refrigerant in the chilling train, where hydrogen is separated out.

In order to investigate the energy efficiency of the ethylene plant, it is necessary to identify the main input and output indicators. For ethylene plant, the factors directly related to production efficiency mainly include: (1) raw materials; (2) fuel and power consumption; (3) products. Namely crude oil (light diesel, stuff naphtha, other stuff), fuel (fuel gas), steam (ultra-high pressure steam, high pressure steam, (production) medium pressure steam, low pressure steam), water (recycled water, industrial water, (production) boil feed water, other water), electricity are the input indicators of ethylene production, while the yields of ethylene and propylene, ethylene production loading rate and are the output indicators [18,19,26,27]. The energy utilization boundary of ethylene product system is described in Fig. 3.

Currently, the common method to express the ethylene plant energy consumption level is to uniformly convert the measurement units of fuel, steam, water and electricity of energy-related parameters into GJ, in accordance with the conversion relationship in Table 3.0.2 and Table 3.0.3 in "Petrochemical design energy consumption calculation method" (SH/T3110-2001) [28], and then summed (SEC = fuel + steam + water + electricity). The method may reflect the energy consumption level of the plant to some extent, but only considering the fuel and power variable, without other parameters related to the ethylene plant energy consumption, such as raw material (light diesel oil, naphtha, etc.), ethylene production, ethylene yield and plant load rate, etc. Therefore, it cannot be used to reveal the effects of the same plant using different materials or different procedures on the plant energy consumption, causing certain limitations in the actual application. To this end it is necessary to utilize the time-series data of all variables related to ethylene plant energy consumption to integrate a value as the energy efficiency benchmarking. Through getting more information of energy consumption level of the plant, it reveals the real reason for affecting this.

This paper selects the monthly production data of 19 ethylene plants in nationwide, during the period 2001 to 2010 by seven major ethylene production technologies as the analysis objects [27]. Firstly, choose the real statistics related to energy efficiency of an ethylene plant during the period 2001 to 2010 (total 120 months) under sequential separation technology as the input data to verify the proposed fusion method. After normalized by Eqs. (4) and (5), the input data are clustered by the proposed fuzzy C-means algorithm. Total 4 classes are obtained, and the clustering centers and the variance results of each variable corresponding to various types clustering centers are shown in Fig. 4 (The dotted circle is the class, and the other circles are the data).

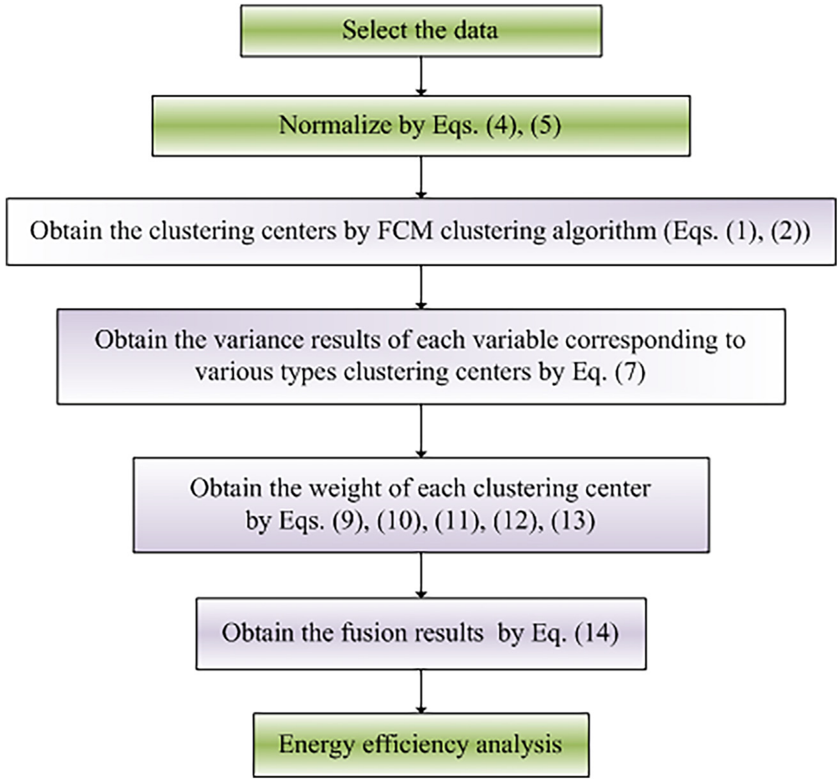

Fig. 1. Flowchart of linear optimization fusion model based on fuzzy C-means.

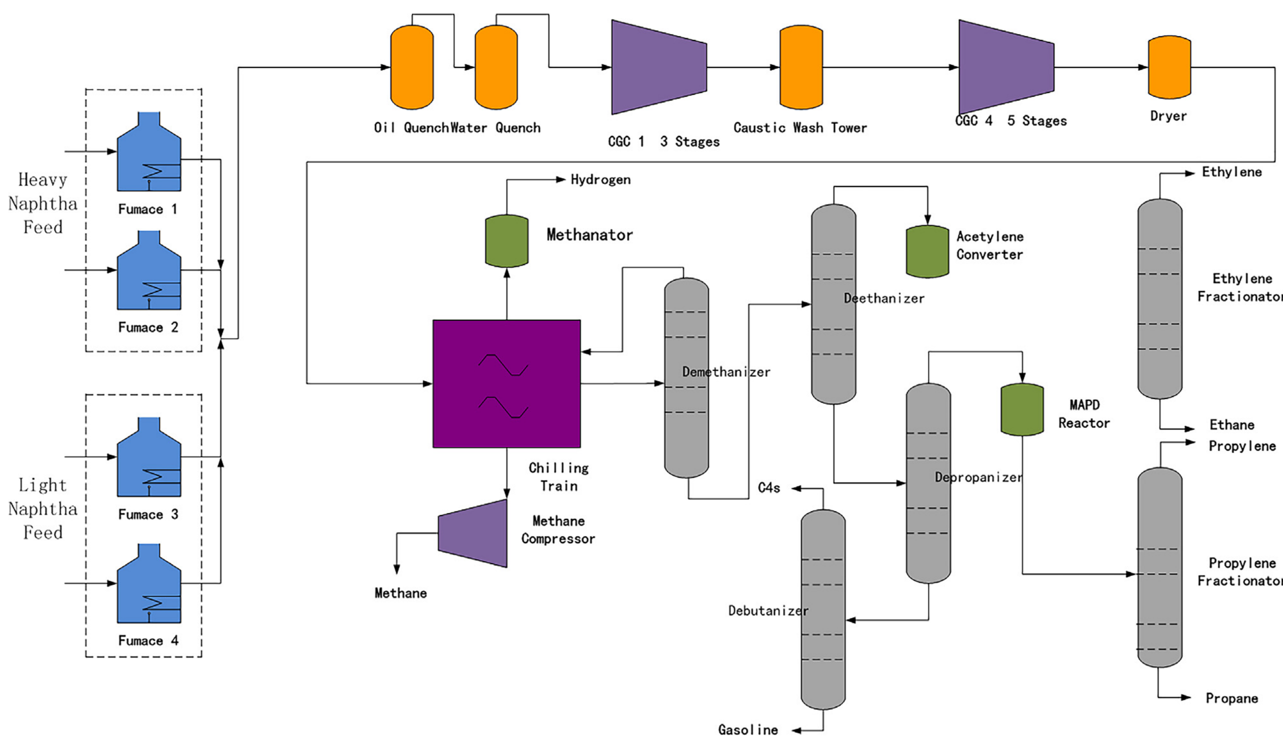

Fig. 2. A typical framework of a typical ethylene plant.

The clustering centers and the mean square errors in Fig. 4 and the input data are substituted into Eqs. (12) and (13), obtaining the weight vector W of various types clustering centers data.

$$W^T = [0.2160\ 0.2992\ 0.1825\ 0.3022] \tag{15}$$

While the weight value obtained from the literature [12] is $W^T = [-0.7631\ 0.0561\ 2.0108\ -0.3039]$, where excessive negative weights would clearly have a bad influence on the optimization fusion results.

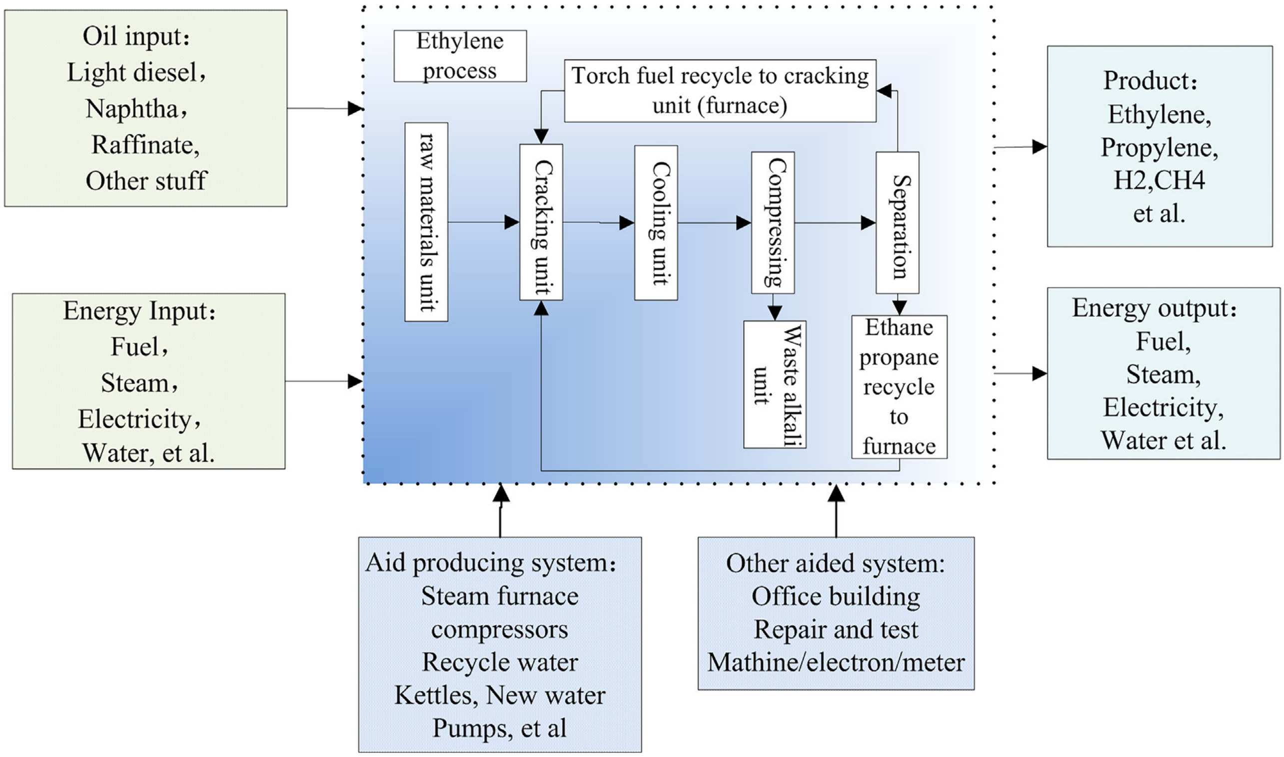

Fig. 3. The energy utilization boundary of ethylene product system.

Table 1
Linear optimal results of time-series data for ethylene plant energy efficiency.

Variable parameters	Class 1	Class 2	Class 3	Class 4	Fusion result (Virtual benchmarking)	
Crude oil	Light diesel/t	644.41	107.14	488.55	267.78	341.38
	Stuff naphtha/t	37189.00	56652.35	37288.40	55567.62	48585.26
	other stuff/t	7427.54	393.21	6912.17	1207.10	3348.76
Ethylene production/t	18968.36	18131.84	19002.46	18644.48	18626.41	
Ethylene yield/%	0.33	0.31	0.33	0.31	0.32	
Fuel gas/t	0.49	0.54	0.49	0.52	0.51	
Ultra-high pressure steam/t	0.40	0.06	0.44	0.42	0.31	
High-pressure steam/t	1.15	0.60	1.13	0.89	0.90	
Medium-pressure steam/t	−0.25	−0.29	−0.24	−0.29	−0.27	
Low-pressure steam/t	−0.27	−0.07	−0.21	−0.09	−0.15	
Recycled water/t	291.32	357.53	279.86	291.28	309.03	
Industrial water/t	0.20	0.20	0.20	0.20	0.20	
Boil feed water/t	2.28	2.53	2.26	2.61	2.45	
Other water/t	0.01	0.00	0.01	0.00	0.01	
electricity/kW h	1.90	2.00	1.89	1.92	1.93	
loading rate/%	104.88	100.20	105.18	102.12	102.70	
SEC/GJ (t ethylene) $^{-1}$	25.89	27.72	25.96	26.94	26.77	
Length of the data	35	22	32	24	113	

Note: The negative values in the table indicates that the variable is negative effect variable.

The weight values of clustering centers in Eq. (15) substituted into Eq. (14), it can obtain the calculation results of linear optimization fusion of ethylene plant energy efficiency related data, and then use Eqs. (4) and (5) anti-normalization, obtaining the fusion results as shown in Table 1 that is the energy efficiency benchmark.

In order to analyze ethylene energy efficiency value better, uniformly convert the units of fuel, steam, water and electricity into GJ, in accordance with the conversion relationship in Table 3.0.2 and Table 3.0.3 in "Petrochemical design energy consumption calculation method" (SH/T3110-2001) [28]. As for the major producing ethylene crude oil naphtha, hydrogenated tail oil, light diesel oil, raffinate oil, carbon-345 and so on, summed as one crude oil index (feedstock) for processing. Select the energy efficiency data of nearly 10 years annual per ton ethylene and the main parameters fusion data related to energy efficiency under two different technologies as shown in Figs. 5 and 6.

As shown in Fig. 5, the ethylene production scales are different under the same technology, but the ethylene yields change little. With the increase of the scale, the propylene yield increases and SEC gradually reduces. Under the ethylene production scale from 150,000 to 200,000 ton, the ethylene yield is the highest and the recycling steam in the production is greater than the input steam. But the amount of the consumed fuel, water and electricity for per ton ethylene production is also the maximum. Therefore, the SEC is the highest. Under the ethylene production scale from 200,000 to 600,000 ton and 600,000 to 800,000 ton, the consumed electric-

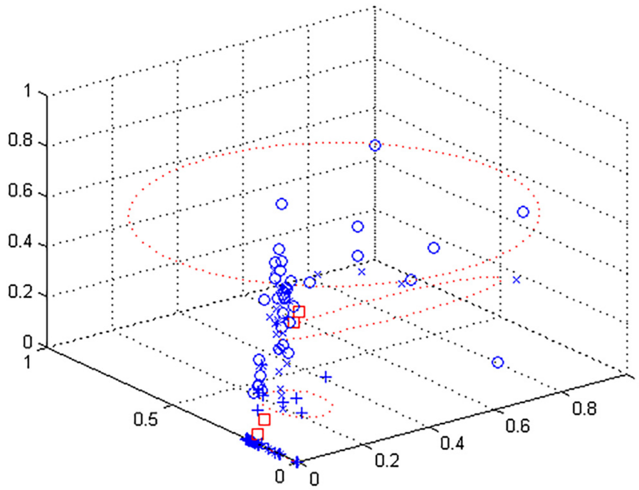

Fig. 4. 3D chart clustering results of time-series data for ethylene plant energy efficiency.

ity for per ton ethylene production is the least and the production load is the largest. Under the ethylene production scale from more than 1 million tons, the consumed steam, fuel, water and electricity for per ton ethylene production (the SEC) are the least. But the required crude oil has increased greatly. Under the technology, different ethylene production plants can reasonably produce ethylene production according to the shortage degree of the raw materials for ethylene production.

Similarly Fig. 6 shows the situations of the consumed crude oil, steam, fuel, water and electricity and their improvement direction of different ethylene production plant under various ethylene production scales with another technique. Figs. 5 and 6 show that, under the two techniques, the annual average energy consumption of the technique in Fig. 5 is significantly lower than the technique in Fig. 6. The main reason is that the SEC of the technique in Fig. 6 is relatively high under the ethylene production scale from 150 to 200 thousand tons, pulling low the average energy consumption. However, under different technologies, the yields of ethylene and propylene are the basically same with the same ethylene production scale. The SEC is less than the normal under the technique in Fig. 5. But the desired crude oil is required more. It is the opposite of the technique in Fig. 6. Without changing the technology, the ethylene production scale can be appropriately adjusted according to the local market price and the supply and the demand relationship

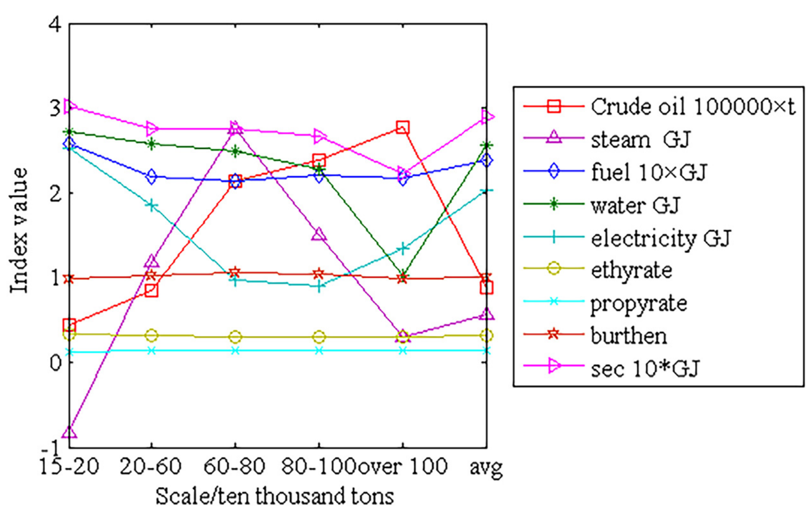

Fig. 5. Energy efficiency chart of annual per ton ethylene with one technology.

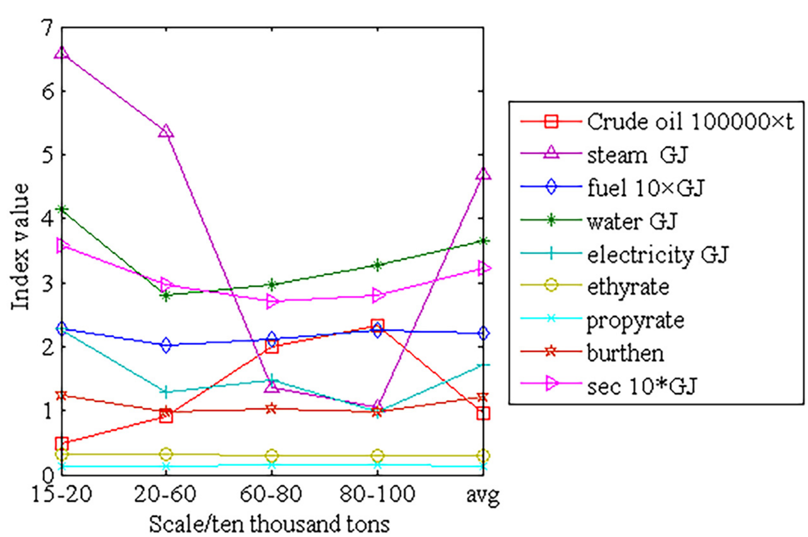

Fig. 6. Energy efficiency chart of annual per ton ethylene with another technology.

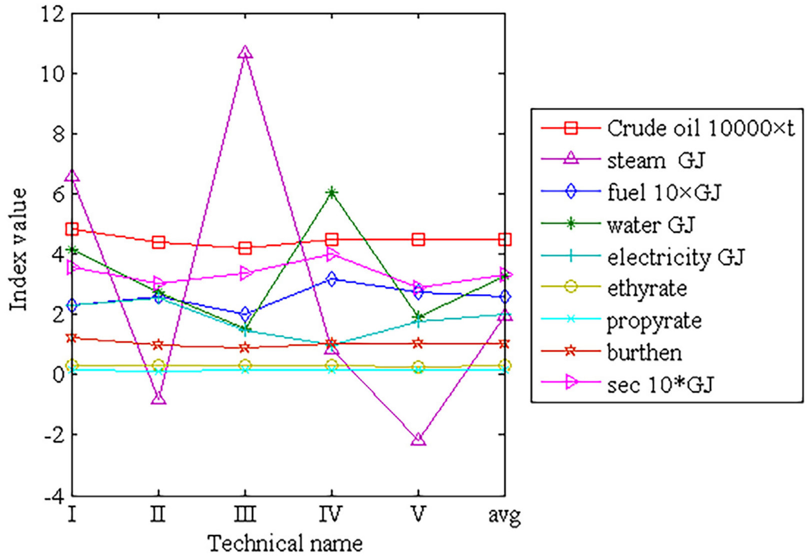

Fig. 7. Energy efficiency chart of annual per ton ethylene with 150,000 to 200,000 ton of ethylene production scale.

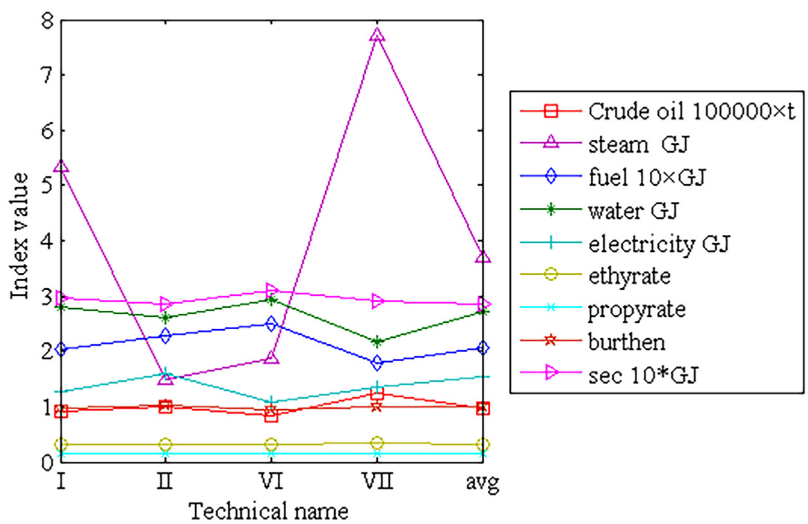

Fig. 8. Energy efficiency chart of annual per ton ethylene with 200,000 to 600,000 ton of ethylene production scale.

of the raw material for ethylene production, and then the optimal production status of ethylene production plants is achieved.

Similarly it can obtain the ethylene energy efficiency and energy efficiency fusion data of nearly 10 years under different ethylene production technologies with two scales, as shown in Figs. 7 and 8.

As shown in Fig. 7, the sums of ethylene yield and propylene yield are similar under different technologies with ethylene production scales from 150,000 to 200,000 ton. However, under technology V, the lowest yield of ethylene and propylene is only 0.4216, and the SEC is lowest too. The steam consumption is negative under technology II and V, indicating produced steam more than the input. The fuel and water consumptions of Technology III are least. But the steam consumption of Technology III is the largest. Technique I has the most crude oil consumption and SEC, as well as the load rate is the highest, leading to the lower level of energy efficiency.

As shown in Fig. 8, the yields of ethylene and propylene are different under different technologies with ethylene production scales from 200,000 to 600,000 ton. The ethylene and propylene yields are the highest under technology VI. But the SEC consumption and the energy consumption of technology VI are the highest. So the production under this technology basically stopped after 2005. However, the ethylene production of others technologies basically remains stable. The load rate basically remains normal. With the comparison of ethylene production scales in Figs. 7 and 8, the larger ethylene production scale with the same technology, the higher the yields of ethylene and propylene and the lower SEC consumption. Although the consumption of crude oil is relatively more, its energy efficiency is relatively high. The reason is that the increase in scale and the decrease in the required energy consumption of ethylene production. The ratios of the ethylene and propylene yields and the consumption of crude oil are higher. In the case of ethylene production scale unchanged, in order to guide the direction of energy efficiency improvements, various ethylene production plants can compare the reason of ethylene production causing energy consumption under different techniques and learn from the local energy supply and demand effect.

By comparing the energy efficiency curves of ethylene production plants with different technologies and different scales, the results can reflect the status of energy consumption and the crude oil consumption and the improvement direction of energy efficiency. Meanwhile, the results can visually reveal the running state and the operational level of each annual plant under different technologies and different scales, accurately identify the main reason for energy consumption, and indicate the main directions to improve energy efficiency.

5. Conclusions

This paper proposes a linear optimization fusion algorithm of the time-series data based on the fuzzy C-means clustering. Application examples in energy efficiency data of ethylene plants verify that it is effective for the large-scale time-series data adopting the fuzzy C-means clustering algorithm to reduce data dimension, exclude abnormal points and obtain the clustering centers. The mean square error of each time-series variable for each clustering center is obtained by the FCM clustering. Based on the above mean square errors, the linear optimization fusion weighted vectors of the multivariate time-series data are calculated to avoid the problem of the negative fusion weights. Then we obtain the reasonable benchmark of energy efficiency values of the ethylene plant by using the proposed linear optimization fusion method. The benchmark combined with curves of energy efficiency can directly reflect the value of multi-dimensional results of fusion and guide ethylene production plants to find the main factors for energy consumption. Meanwhile, the proposed method indicates the direction of improving energy efficiency under different technologies and different scales.

In our further studies, we will take the effect of economic development, and environmental planning on the energy efficiency of ethylene product plants into account. Moreover, we will study artificial neural network integrated DEA with variable return to scale to analyze the input-output energy measuring of ethylene product process, and to compare with the current work. Furthermore, the proposed method can be widely application in energy efficiency evaluation of other large-scale process industry.

Acknowledgments

This research was partly funded by National Natural Science Foundation of China (61533003, 61603025) and Natural Science Foundation of Beijing, China (4162045).

References

[1] G.F. Ma, Y.H. Xu, X. Guo, Ethylene business review of china petrochemical in 2014, Ethyl. Ind. 27 (1) (2015) 1–5.
[2] L.J. Zhang, J. Hu, Ethylene business review of China petroleum in 2014, Ethyl. Ind. 27 (1) (2015) 6–10.
[3] R. Tao, P. Martin, Olefins from conventional and heavy feedstocks, Energy 31 (2006) 425–451.
[4] J.A. Baileya, R. Gordona, D. Burtonb, E.K. Yiridoe, Energy conservation on nova scotia farms: baseline energy data, Energy 33 (2008) 1144–1154.
[5] K. Hirata, Energy saving for ethylene process by advanced process integration, Chem. Eng. Trans. 25 (2011) 81–86.
[6] Z.Q. Geng, Z. Wang, Q.X. Zhu, Y.M. Han, Multi-objective operation optimization of ethylene cracking furnace based on AMOPSO algorithm, Chem. Eng. Sci. 153 (2016) 21–33.
[7] M. Yousefi, A.N. Darus, M. Yousefi, D. Hooshyar, Multi-stage thermal-economical optimization of compact heat exchangers: a new evolutionary-based design approach for real-world problems, Appl. Therm. Eng. 83 (2015) 71–80.
[8] J. Fu, Q. Xu, Simultaneous study on energy consumption and emission generation for an ethylene plant under different start-up strategies, Comput. Chem. Eng. 56 (2013) 68–79.
[9] E. Keogh, M. Pazzani, An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback, in: Proceedings of the 4rd International Conference of Knowledge Discovery and Data Mining, AAAI Press, 1998, pp. 239–241.
[10] Y.J. Weng, Z.Y. Zhu, Novel algorithm for time series data mining based on dynamic time warping, Comput. Simul. 21 (3) (2004) 37–40.
[11] J. Zhou, Z.S. Wang, F.Q. Zhou, The theory of multi-senor system data fusion based on linear least square estimation, J. Astronaut. 24 (4) (2003) 364–367.
[12] Q.X. Zhu, X.Y. Shi, X.B. Gu, Q. Tian, Time-series data fusion and its application to energy efficiency value for ethylene plants, J. Chem. Ind. Eng. 61 (10) (2010) 2620–2626.
[13] Z.Q. Geng, X.Y. Shi, X.B. Gu, Q.X. Zhu, Hierarchical linear optimal fusion algorithm and its application in ethylene energy consumption indices acquisition, J. Chem. Ind. Eng. 61 (2010) 2056–2060.
[14] G.F. Bin, X.J. Li, B.S. Dhillon, W.W. Chu, Quantitative system evaluation method for equipment state using fuzzy and analytic hierarchy process, Syst. Eng.-Theory Pract. 30 (4) (2010) 744–750.
[15] Y.M. Han, Z.Q. Geng, Liu. Qiyu, Energy efficiency evaluation based on data envelopment analysis integrated analytic hierarchy process in ethylene production, Chin. J. Chem. Eng. 22 (12) (2014) 1279–1284.
[16] Y.M. Han, Z.Q. Geng, X.B. Gu, Z. Wang, Performance analysis of china ethylene plants by measuring Malmquist production efficiency based on an improved data envelopment analysis cross-model, Ind. Eng. Chem. Res. 54 (2015) 272–284.
[17] Y.M. Han, Z.Q. Geng, Z. Wang, P. Mu, Performance analysis and optimal temperature selection of ethylene cracking furnaces: a data envelopment analysis cross-model integrated analytic hierarchy process, J. Anal. Appl. Pyrol. 122 (2016) 35–44.
[18] Z.Q. Geng, Q.X. Zhu, X.B. Gu, Dependent function analytic hierarchy process model for energy efficiency virtual benchmark and its applications in ethylene equipments, J. Chem. Ind. Eng. 62 (2011) 2372–2377.
[19] Z.Q. Geng, Y.M. Han, C.P. Yu, Energy efficiency evaluation of ethylene product system based on density clustering data envelopment analysis model, Adv. Sci. Lett. 5 (2012) 1–7.
[20] C.T. Yiakopoulos, K.C. Gryllias, I.A. Antoniadis, Rolling element bearing fault detection in industrial environments based on a K-means clustering approach, Expert Syst. Appl. 38 (2011) 2888–2911.
[21] M. Erisoglu, N. Calis, S. Sakallioglu, A new algorithm for initial cluster centers in k-means algorithm, Pattern Recognit. Lett. 32 (2011) 1701–1705.
[22] Y.F. Xu, C.M. Chen, Y.Q. Xu, An improved clustering algorithm for K2-means, Comput. Syst. Appl. 25 (3) (2008) 27–31.
[23] J. Yu, Q.S. Cheng, H.K. Huang, Analysis of the weighting exponent in the FCM, IEEE Trans. Syst. Man Cybern. Part B – Cybern. 34 (2004) 634–639.
[24] M.J. Fadili, S. Ruan, D. Bloyet, B. Mayoyer, On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series, Med. Image Anal. 5 (2001) 55–67.
[25] X.B. Gu, Q.X. Zhu, Fuzzy multi-attribute decision-marking method based on eigenvector of fuzzy attribute evaluation space, Decis. Support Syst. 41 (2006) 400–410.
[26] X.L. Wang, The a dynamic benchmarking method for energy consumption of ethylene production process, Comput. Appl. Chem. 27 (9) (2010) 166–1170.
[27] Z.Q. Geng, Y.M. Han, X.B. Gu, Q.X. Zhu, Energy efficiency estimation based on data fusion strategy: case study of ethylene product industry, Ind. Eng. Chem. Res. 51 (2012) 8526–8534.
[28] Calculation method for energy consumption in petrochemical engineering design. SH/T3110-2001 (2002).