A Novel MDFA-MKECA Method With Application to Industrial Batch Process Monitoring

Yinghua Yang, Xiang Shi, Xiaozhi Liu, and Hongru Li

Abstract—For the complex batch process with characteristics of unequal batch data length, a novel data-driven batch process monitoring method is proposed based on mixed data features analysis and multi-way kernel entropy component analysis (MDFA-MKECA) in this paper. Combining the mechanistic knowledge, different mixed data features of each batch including statistical and thermodynamics entropy features, are extracted to finish data pre-processing. After that, MKECA is applied to reduce data dimensionality and finally establish a monitoring model. The proposed method is applied to a reheating furnace industry process, and the experimental results demonstrate that the MDFA-MKECA method can reduce the calculated amount and effectively provide on-line monitoring of the batch process.

Index Terms—MDFA, MKECA, process monitoring, reheating furnace, statistical features, thermodynamics entropy feature.

I. INTRODUCTION

BATCH and semi-batch processes, as the traditional industrial processes, have been generally used in the chemical, food, biochemical, and semiconductor industries [1]. In order to ensure the safety of these industrial batch processes and improve the final quality of products, on-line process monitoring is becoming increasingly important. Multivariate statistical process monitoring (MSPM) [2] is a powerful tool for the comprehensive monitoring of industry processes and detection of abnormal operation. It has been widely applied to both continuous and batch processes with many successful applications. For continuous processes, principal component analysis (PCA) [3] and partial least squares (PLS) [4], as two well-known typical MSPM methods, have been utilized to monitor performance and quality. Batch process data typically has a three-dimensional data structure. However, PCA and PLS can only handle two-dimensional data matrix, so they are not suitable. Therefore, multi-way principal component analysis (MPCA) [5], multi-way partial least square (MPLS) [6], and multi-way

The authors are with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: yhyang@mail.neu.edu.cn; 438798694@qq.com; liuxiaozhi@mail.neu.edu.cn; lihongru@mail.neu.edu.cn).

independent component analysis (MICA) [7] are proposed to adapted to batch processes. In order to tackle process nonlinearity, the kernel functions are introduced and then multi-way kernel principal component analysis (MKPCA) [8], MKICA [9], and MKPLS [10] are developed as an extension of regular methods. Considering that more data information can be revealed by information entropy, recently, the multi-way kernel entropy component analysis (MKECA) method has been proposed [11]. It is a new method of data transformation and dimensionality reduction, which chooses the best principal component vectors according to the maximal Renyi entropy rather than judging by the top eigenvalues and eigenvectors of the kernel matrix simply. The MKECA method provides a more effective way for batch process monitoring.

A real industrial process, such as a reheating furnace, has a heavy and unequal batch, plentiful process measurements, nonlinear behavior, and other complex characteristics. For these industries processes, the MKECA monitoring method is not useful since its basis are all batches with equal data length. Another disadvantage is that as a second-order method, the MKECA method loses sight of data information included in higher-order representations, such as non-Gaussian, which are universal characteristics of industrial data. Additionally, the calculated amount is also enormous due to a large amount of measurements and projections based on the kernel matrix, which has a negative impact to the on-line monitoring. Hence, to overcome the above disadvantages, a data pre-processing method before MKECA is needed. In the literatures, data pre-processing methods for batch data of unequal length have been developed, including "Minimum length", "Maximum length" [12], and "Indicator variable" [13]. However, these methods are not efficient enough in the real industrial processes because of the many operational stages and uncertain length of the batch data. Alternately, the dynamic time warping (DTW) algorithm [14], as a time-series similarity measure which minimizes the effects of shifting and distortion, is developed to solve the problem of unequal batch data. In DTW, the batch data length is made the same by conversion, expansion or compression of some local data fragments. However, this method is only suitable for processes whose batch data length have little difference. Once there are great differences, an unrealistic correspondence will occur.

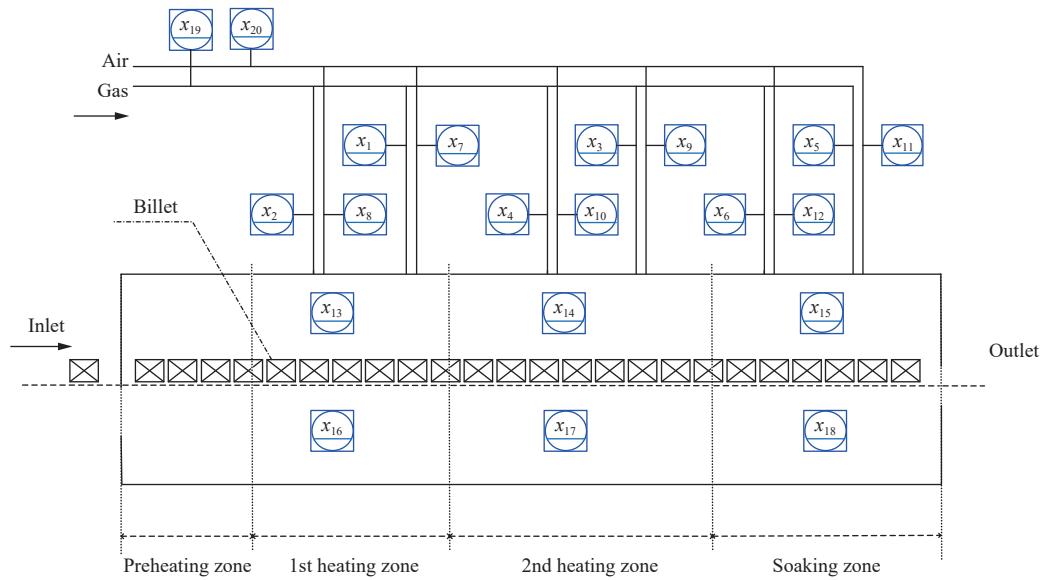As an important means of representing data information,

Fig. 1.   The structural representation of walking beam reheating furnace.

data feature extraction is widely used in various data processing [15], [16]. In this paper, based on the integration of mixed data features analysis and multi-way kernel entropy component analysis, a novel batch process monitoring method named MDFA-MKECA is proposed to obtain better monitoring performance. This method consists of two phases: data pre-processing (by MDFA) and process monitoring (by MKECA). The main idea of pre-processing is that combined with mechanistic knowledge, various data features involving statistics features and thermodynamics entropy features are extracted and calculated. In addition to reduce the size of data and achieve the purpose of equal length for different batches, complex characteristics of data can also be addressed effectively by this step. In the second phase, MKECA is used to reduce data dimensionality and establish a monitoring model. Three-dimensional feature data is unfolded and then KECA is applied to choose the best principal component vectors, which ensures that the data information is lost slightly during the dimensionality reduction. The MDFA-MKECA method exerts the advantage of MDFA in processing unequal batch data, reducing data size and tackling the complex characteristics of industrial process, which makes up for the lack of MKECA in data pre-processing. Finally, it is applied in the reheating furnace process analysis and real-time monitoring with great performance. This paper is organized as follows. A reheating furnace industry process is introduced in Section II. Section III focuses on the selection and extraction procedure of mixed data features with engineering and mechanistic knowledge. In Section IV, the further discussion of the proposed MDFA-MKECA is described. The simulation results and discussion are given in Section V. Finally, we conclude the paper in Section VI.

II.   AN INDUSTRIAL CASE RESEARCH BACKGROUND

As a typical industry process with the above complex characteristics, a reheating furnace is an important piece of equipment for the hot rolling process of iron & steel and nonferrous metal, where the billets are heated to a preset temperature before entering the rolling mill [17]. A structural representation of the furnace is shown in Fig. 1. In the reheating furnace, the billets are fed to the preheating zone, and then moved through three combustion zones (including two heating zones and a soaking zone) sequentially to the outlet by the walking beam. Throughout the duration of billet heating, the gas and air flow is controlled so that the billets reach the preset temperature when arriving at the outlet.

The accuracy and the uniformity of the billet outlet temperature are two important indexes to evaluate the quality of reheating furnace running state. As the temperature of the moving billets cannot be measured directly in the furnace, the heat exchange model [18] has been a major and vital means to monitor the distribution of the billet temperature. However, more than 90% heat exchange is based on heat radiation, and radiation coefficients are susceptible to various factors in the furnace, i.e, the heat exchange coefficient is often affected by the changed operating state of furnace. Therefore, the model calculation result often does not correspond with the real billet temperature. Once the production state deviates from the normal condition, for example if the value of the air-fuel ratio, fuel flow and production rate change, that can cause the calculated temperature to be higher (or lower) than the actual temperature, and bad reheating quality of billets (or waste of fuel) will occur as a result. The process monitoring is mainly used to detect the deviation level between the running state and standard state of the heating process, which helps to adjust the control parameters and correct the deviation in time. Furthermore, good control performance can be achieved and the heating quality is also improved.

In the reheating furnace discussed above, each billet is moved through the same process in four zones, which in turn can be regarded as one batch. However, the variation in production rhythm for various billets results in different

heating times and heat exchange. Hence, from the point of view of the billets, the reheating furnace is regarded as a typical complex industrial batch process to study in this paper.

There are 20 measured process variables x_1–x_{20} involved in furnace operation as shown in Table I. Beside the gas and air pressure of main pipes, each combustion zone also has 6 variables including gas flow, air flow and temperature of upper and lower part, respectively. The major characteristics of process variables are as follows:

1) Nonlinearity: The heat exchange within the furnace is due to heat release from the combustion of fuel and the absorption of heat by the billet mainly, and these heat transfer modes including radiation heat transfer and convection heat transfer have typical non-linear characteristics.

2) Heavy Unequal Batch Data: Each billet is regarded as a unit and its heating process is defined as one batch. Influenced by the variational production rhythm, the heating time is not consistent for different batches. For example, although the standard heating time of billet is about 3.5 hours, some of billets' heating time will be 4 or 6 hours when the rolling rhythm is changed. Hence, for the fixed time sampling data, the length of each batch may not be equal.

3) A Large Amount of Process Measurements: The data sample interval is ten seconds while billet heating time is 3.5 to 6 hours and up to 40 billets are in the reheating furnace simultaneously, which results in a large amount of data being recorded, enormous calculated resources are needed to monitor temperature profile of all the billets in reheating furnace.

TABLE I
REHEATING FURNACE PROCESS VARIABLES DESCRIPTION

Serial number	Variable name	The description of variable
1, 3, 5	x_1, x_3, x_5	Gas flow of upper part in three combustion zones
2, 4, 6	x_2, x_4, x_6	Gas flow of lower part in three combustion zones
7, 9, 11	x_7, x_9, x_{11}	Air flow of upper part in three combustion zones
8, 10, 12	x_8, x_{10}, x_{12}	Air flow of lower part in three combustion zones
13, 15, 17	x_{13}, x_{15}, x_{17}	Upper part temperature of three combustion zones
14, 16, 18	x_{14}, x_{16}, x_{18}	Lower part temperature of three combustion zones
19	x_{19}	Gas pressure
20	x_{20}	Air pressure

III. DATA PRE-PROCESSING BASED ON MIXED DATA FEATURES ANALYSIS

According to the above process analysis, there are heavy unequal batches, a large amount of process measurements, nonlinear behavior and other characteristics in the reheating furnace industry that results in its complex mechanism and inner structure. Consequently, general batch process monitoring methods such as MKECA cannot be applied directly to the original data. Several steps of data pre-processing such as trajectory warping and centralized criterion are needed, because of plentiful batch data and the existence of kernel matrix, which leads to extensive calculation during

model building.

In this section, we propose a novel data pre-processing method based on mixed data features analysis (MDFA). This method can not only eliminate the data pre-processing steps mentioned above but allows us to acquire more data information. The main goal of MDFA is to analyze data information included in statistic and thermodynamic entropy features, and then select and extract different data features of each batch with mechanistic knowledge. The features selected from different batches are consistent, which allows different batch data to be compressed into the same feature row vector or matrix so that the unequal batch problem can be solved effectively. Meanwhile, the size of the process data is reduced significantly because the number of mixed data features is much smaller than the original data. Additionally, as the statistic features consist of first-order, second-order and high-order statistics, more data information can be captured and the complexity like nonlinearity can also be addressed effectively.

The mixed data features D is defined as

$$D = [D_{SF}, D_{TEF}] \tag{1}$$

where D_{SF} and D_{TEF} denote statistics features and thermodynamics entropy features, respectively.

A. Statistics Features Extraction

Statistical pattern analysis (SPA), a multivariate statistical monitoring framework, is proposed by [19] and applied to extract statistical features of original data. Different process behavior is represented by various statistics including first-order, second-order, and high-order statistics of the process variables. Therefore, the abnormalities of the system are easily captured since the statistical distribution of processes under abnormal conditions would result in obvious information for process monitoring. The following part shows that the statistic characteristics of a batch trajectory are extracted by calculating various statistics.

X_p is used to denote the pth batch of process measurements, shown below:

$$X_p = [x_1, x_2, \ldots, x_m]_{n \times m}$$
$$= \begin{bmatrix} x_1(1) & x_2(1) & \ldots & x_m(1) \\ x_1(2) & x_2(2) & \ldots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(n) & x_2(n) & \ldots & x_m(n) \end{bmatrix} \tag{2}$$

where m is the number of variables measured and n is the batch duration. For different batches records, m is same while n takes different values.

In general, three groups of batch statistics are included in a statistics feature

$$D_{SF} = [\mu|\Sigma|\Xi] \tag{3}$$

where $\mu = [\mu_i]_{1 \times m}$ denotes the first-order statistics, namely,

$$\mu = [\mu_1, \mu_2, \ldots \mu_m]$$

containing the means of all process variables, the single variable mean calculated as follow:

$$\mu_i = \frac{1}{n} \sum_{l=1}^{n} x_i(l). \tag{4}$$

The covariance is selected as second-order statistics, which is defined as

$$\text{cov}(x_i, x_j) = \frac{1}{n-1} \sum_{l=1}^{n} (x_i(l) - \mu_i)(x_j(l) - \mu_j). \tag{5}$$

$C_{i,j}$ is used to denote $\text{cov}(x_i, x_j)$, for simplicity, which represents the covariance between two random variables (x_i, x_j). All the covariances are arranged in the following order to form $\Sigma = [\text{cov}(x_i, x_j)]_{1 \times \frac{m(m+1)}{2}}$, namely:

$$\Sigma = [C_{1,1} C_{1,2} \ldots C_{1,m} C_{2,2} \ldots C_{2,m} \cdots C_{m-1,m} C_{m,m}]. \tag{6}$$

The higher-order statistics $\Xi = \begin{bmatrix} [s_i]_{1 \times m} & [k_i]_{1 \times m} \end{bmatrix}$ include the skewnesses and kurtoses of all process variables.

$$s_i = \frac{E\left[(x_i - \mu_i)^3\right]}{E\left[(x_i - \mu_i)^2\right]^{3/2}} \tag{7}$$

$$k_i = \frac{E[(x_i - \mu_i)^4]}{E[(x_i - \mu_i)^2]^2} - 3. \tag{8}$$

B. Thermodynamics Entropy Feature Extraction

The heating process for the billet is a thermodynamic process where heat is always transferred and exchanged continually. It is tied closely to the running state of the reheating furnace. Statistic features by themselves are not sufficient to describe energy transfer, hence, thermodynamic entropy features extraction is applied to represent all the production states and heat exchange experienced by the billet.

Entropy is an important concept of thermodynamics [20]. It was created by German physicist Rudolf Clausius in the 1850s and 1860s and used to interpret the second law of thermodynamics. In the viewpoint of thermodynamics, according to the Carnot cycle, an isolated system's entropy never decreases.

The quantitative and macroscopical representation of the irreversibility of the system is an increase in entropy. Similarly, the microscopic quantitative description is interpreted by an increase in the number of microscopic states or the thermodynamic probability of system. There are certain Boltzmann relations between the number of microscopic states and the entropy of the system.

$$S = k \ln w. \tag{9}$$

It means that the entropy of system S equals the natural logarithm of number of states w, multiplied by the Boltzmann constant k. The expression denotes that with a higher number of microscopic states, the disorder and chaos will be higher. Namely, the entropy is a measure of the disorder for system molecular thermal motion and the chaos of system.

For a reheating furnace, the heat exchange of system and the billet temperature rise are macroscopic performance of the heating process. The nature of the heating process, namely a microcosmic representation, is the disorderly movement of the

thermal molecules associated with the production state. Hence, according to the above expression, thermodynamic entropy can be a good reflection of running state changes of the reheating furnace. However, it is difficult to extract and calculate thermodynamics entropy E_t, directly in the real production due to its complex internal mechanisms and energy exchange. Hence, information entropy E_i is introduced and applied in computing the thermodynamic entropy. Information entropy can be regarded as a measure of disorder and chaos, which explains the uncertainty of the state of motion for a system. Modern information theory has proved that mathematical quantitative relations between thermodynamic entropy E_t and information entropy E_i are

$$E_t = k \ln 2 E_i. \tag{10}$$

That is to say, the $k \ln 2$ thermodynamics entropy of the system is increased at least enough to compensate when 1 bit information is received by the system. A classic type of information entropy is Renyi entropy and it can be seen in the first part of Section IV. The thermodynamics entropy features D_{TEF} of each batch data can be extracted by two steps: the Renyi entropy E_i of each batch data is calculated at first, and then this information entropy is transformed into thermodynamics entropy E_t by (10).

C. Data Pre-Processing Method Based on Mixed Data Features Analysis

On the basis of the above analysis, mixed data features (MDF) including statistics and thermodynamics entropy are acquired and the dimension of the MDF row vector D is $(m^2 + 7m + 2)/2$.

$$D = [D_{SF}, D_{TEF}]$$
$$= \left[[\mu_i]_{1 \times m}, [\text{cov}(x_i, x_j)]_{1 \times \frac{m(m+1)}{2}}, [s_i]_{1 \times m}, [k_i]_{1 \times m}, E_t \right]. \tag{11}$$

The heating process consists of four combustion zones. Their flow and pressure characteristics are not exactly the same because of different combustion characteristics and target temperatures. Consequently, the monitoring results might be inaccurate if a MDF is obtained by extracting features of a whole batch. That is to say, it is not reasonable to regard batch data as a whole. Engineering and mechanistic knowledge are applied to divide the batch data into different windows of measurements before extracting MDFs in this section. Mechanistic knowledge is a description of inner mechanisms and phenomena of a system, which can provide much more information on the process and is of paramount importance to process monitoring.

The reheating furnace is divided into four zones in Section II and there are variables measurements only in the heating and soaking zones. Therefore, an original batch data should be denoted by a MDF matrix D_{whole} including three MDF row vectors corresponding to 1st heating, 2nd heating, and soaking zones measurements, respectively. Then these matrices are used on behalf of original data and involved in the process monitoring later.
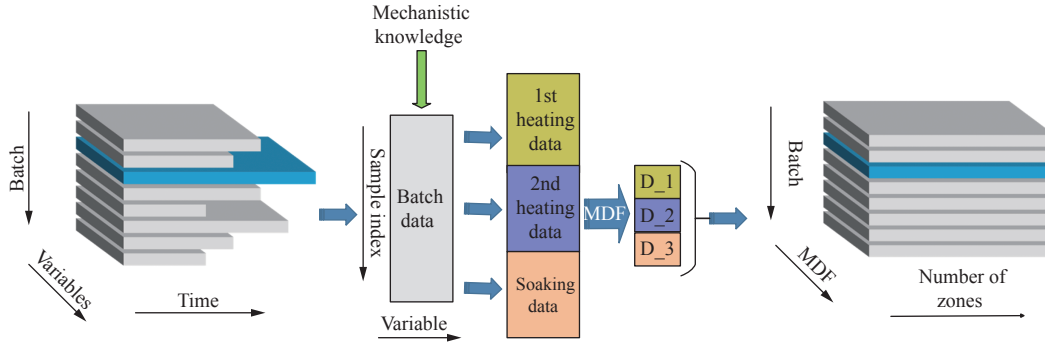
Fig. 2.   The schematic plot of the MDFA framework for data pre-processing.

$$D_{\text{whole}} = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} D_{SF1} & D_{TEF1} \\ D_{SF2} & D_{TEF2} \\ D_{SF3} & D_{TEF3} \end{bmatrix}. \tag{12}$$

It is worth noting that, the dimension of the MDF matrix is $3 \times [(m^2 + 7m + 2)/2]$, and is smaller compared to the original data matrix $(n \times m)$, as the batch duration n is usually much larger than the number of variables m. Hence, the mixed data feature extraction method has great performance on reducing the size of data. In addition, as the existence of higher-order statistics, the complexity can also be addressed effectively.

Fig. 2 shows the detailed steps involved in the data pre-processing method based on mixed data features analysis. Firstly, original batch data is divided into three data areas corresponding to three combustion zones with mechanistic knowledge. And then, a MDF matrix can be acquired by calculating various data features. As the MDF matrices are equal in dimension for different batches, a new MDF three-dimensional data feature is obtained. The data-driven MDFA model is combined with engineering and mechanistic knowledge, which has a great advantages in reflecting the real process status, making the MDFA model more accuracy and reliability. Moreover, monitoring statistics from each batch in each combustion zone can be calculated by introducing mechanistic knowledge, which can capture more specific abnormalities of the running state.

IV.   MDFA-MKECA FOR PROCESS MONITORING

A data pre-processing method based on MDFA has been discussed in Section III. Thereafter, the proposed MDFA-MKECA method and its application for process monitoring are needed to discuss and display further.

A. Multi-Way Kernel Entropy Component Analysis

After the data pre-processing based on MDFA, the original batch data with an unequal length has been transformed into MDF data with equal batch. MKECA can be used for process monitoring directly.

The data after pre-processing is inputted into a three-dimensional matrix $X_{B \times F \times Z}$, where B is the number of batches, F is the number of mixed data features, and Z is the number of combustion zones. As a continuous process monitoring method, KECA cannot be applied in the three-dimensional array directly, so $X_{B \times F \times Z}$ should be unfolded into two-dimensional matrix. The AT approach [21] is the most frequently used unfolding procedure, because it combines with the advantages of batch-wise unfolding [2] and variable-wise unfolding [22] and thus, can express more process information. After the AT approach, the unfolded data matrix X is acquired and kernel entropy component analysis (KECA) will be performed.

From an information theory view, combing Renyi entropy with the kernel method, Jenssen proposed a novel method called KECA [23]. It has the ability to retain the main information of the data structure and is good in nonlinear data-processing.

The Renyi quadratic entropy is given by

$$H(p) = -\log \int p^2(x)dx \tag{13}$$

where $p(x)$ is the probability density function of the data set $X = [x_1, ..., x_N]$. Since the logarithm is a monotonic function, alternatively, one expression can be used

$$V(p) = \int p^2(x)dx. \tag{14}$$

In order to estimate $V(p)$, and hence $H(p)$, a Parzen window density estimator is invoked. The Parzen window is a non-parametric density estimation method. Using the Parzen window, the probability density estimation is given below:

$$\hat{p}(x) = \frac{1}{N} \sum_{l=1}^{N} K_\sigma(x_i, x_j) \tag{15}$$

where $K_\sigma(x_i, x_j)$ is the Parzen window or kernel centered at x_i, and parameter σ is the kernel size. Using the sample mean approximation of the expectation operator, we get

$$\begin{aligned} \hat{V}(p) &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} K_\sigma(x_i, x_j) \\ &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_\sigma(x_i, x_j) = \frac{1}{N^2} 1^T K 1. \end{aligned} \tag{16}$$

Here, each element of the $N \times N$ kernel matrix K equals $K_\sigma(x_i, x_j)$, and 1 is $(N \times 1)$ column vector of ones. The Renyi entropy estimator may be expressed in terms of the eigenvalues and eigenvectors of the kernel matrix, which may be decomposed as $K = E D_\lambda E^T$, where D_λ is a diagonal matrix storing the eigenvalues $\lambda_1, ..., \lambda_N$ and E is a matrix with the corresponding eigenvectors $\alpha_1, ..., \alpha_N$ as columns. Rewriting

(16), we have

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^{N} (\sqrt{1_i} \alpha_i^T 1)^2 \qquad (17)$$

entropy estimate is contributed to each term in (17), which means that it is contributed to more, by certain eigenvalues and eigenvectors. The eigenvalues and eigenvectors selected are the first l that contribute the most to the entropy estimate in KECA.

B. Process Monitoring Based on MDFA-MKECA

According to the above analysis, in this section, we integrate improved mixed data features analysis and multi-way kernel entropy component analysis to develop a MDFA-MKECA for process monitoring. This strategy is used to construct various mixed data feature sets as the substitutes of process variables and then apply MKECA between training and testing datasets.

The process monitoring using MDFA-MKECA has two phases: off-line modeling and on-line monitoring. Details are given as follows.

1) Off-Line Modeling

a) Firstly, some normal operating original batches data whose lengths are different are selected as training data $X(B \times m \times n)$, where B is the number of batches, m is the number of process variables, and n is the sampling time, as shown in Fig. 2. Each piece of batch data, using the pth batch $X_p(m \times n)$ as an example, is divided into three data areas $X_{p1}(m \times n_1)$, $X_{p2}(m \times n_2)$, and $X_{p3}(m \times n_3)$, corresponding to 1st heating zone, 2nd heating zone, and the soaking zone with mechanistic knowledge;

b) According to (4)–(8) and (10), three data areas are transformed into corresponding three MDF row vectors D_{p1}, D_{p2}, D_{p3} with a dimensional of d. The MDF matrix D_p can then be acquired, where $d = (m^2 + 7m + 2)/2$;

$$D_p = \begin{bmatrix} D_{p1} \\ D_{p2} \\ D_{p3} \end{bmatrix} = \begin{bmatrix} D_{SF1} & D_{TEF1} \\ D_{SF2} & D_{TEF2} \\ D_{SF3} & D_{TEF3} \end{bmatrix}. \qquad (18)$$

c) The new three-dimensional training data $X_M(B \times d \times 3)$ is obtained and composed of B different MDF matrixes with equal dimensions. Then it is unfolded into a two-dimensional matrix $X'_M(3B \times d)$ using the AT approach.

d) Select a radical basis kernel function and parameter σ

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2). \qquad (19)$$

Then, the kernel matrix K can be obtained and applied to the eigen-decomposition $K = E D_\lambda E^T$.

e) According to (16), Renyi entropy which corresponds to each eigenvalue can be estimated and then l eigenvectors are selected according to the contribution towards the entropy estimate.

f) Calculate the kernel feature space data points $\phi_{eca} = D_l^{-1/2} E_l^T$ and score vectors $t = K \times E$.

g) Calculate squared prediction error (SPE) statistics Q and statistic T^2 of the training data

$$Q = \sum_{s=1}^{d} t_s^2 - \sum_{s=1}^{l} t_s^2 \qquad (20)$$

$$T^2 = [t_1, t_2, ..., t_l] \Lambda^{-1} [t_1, t_2, ..., t_l]^T \qquad (21)$$

where Λ^{-1} is the inverse matrix of the covariance of the principal component matrix containing l extracted principal components. The control limits T^2 of and Q are $T^{2(c)}$ and $Q^{(c)}$

$$Q^{(c)} \sim g \chi^2_{h,\alpha} \qquad (22)$$

where $g = v/2m$, $h = 2m^2/v$, α is confidence level, m, v are the mean and variance of Q statistics obtained from the training batches.

$$T^{2(c)} \sim \frac{l(B^2 - 1)}{B(B - l)} F_{l, B-l, \alpha} \qquad (23)$$

where α is confidence level, B is the number of batches and obey the F distribution whose degrees of freedom is l and condition is $B - l$.

2) On-Line Process Monitoring

a) Collect the real-time monitoring new data x_{new}, similar to training data, the MDF matrix D_{new} of new data can be calculated.

b) Afterwards, score vectors $\hat{t} = \hat{K} \times E$ are obtained and the Q and T^2 statistic are calculated according to (20) and (21).

$$Q_{new} = \sum_{s=1}^{d} \hat{t}_s^2 - \sum_{s=1}^{l} \hat{t}_s^2 \qquad (24)$$

$$T^2_{new} = [\hat{t}_1, \hat{t}_2, ..., \hat{t}_l] \hat{\Lambda}^{-1} [\hat{t}_1, \hat{t}_2, ..., \hat{t}_l]^T. \qquad (25)$$

c) Making a comparison between the statistics of the testing MDF matrixes and control limit $T^{2(c)}$ and $Q^{(c)}$ calculated in (22) and (23). If statistics are below the control limit, it is classified as a normal batch; otherwise, it is classified as an abnormal batch.

V. EXPERIMENT AND MONITORING RESULTS

To prove the effectiveness of the proposed method in this paper, some simulation experiments are designed based on real production data provided by a steel mill. The whole data set includes 70 billet batches which consist of 37 batches of data under normal working conditions and 33 batches of data under abnormal working conditions. The batch records are collected by the control system at ten second intervals. The process variables are shown in Table I.

In our experiments, 15 normal batches are used to be training data, and 55 batches of data including 22 batches of data under normal working conditions and 33 batches of data under abnormal working conditions are used as testing data. The original batch includes 5600–9000 pieces of data while there are only 183 pieces of data after data pre-processing, which fully demonstrates the effectiveness of MDFA on reducing the size of the data. The abnormal working conditions are specifically divided into mild abnormalities and heavy abnormalities in testing data, which corresponds to the small and big deviations of outlet temperature between model outputs and measurements in field. In process monitoring, if

the statistic of a batch exceeds control limits of two or three combustion zones, the batch is regarded as a heavy abnormal batch. Analogously, if the overrun occurs only in one or none of the combustion zone, the batch is defined as a mild abnormal or normal batch. Hence, the detection result for a batch will be reasonable when the deviation of the outlet temperature is consistent with the monitoring results.

The monitoring performances of every testing batch in three combustion zones are shown in Fig. 3, respectively. In the monitoring charts, Fig. 3(a) shows the normal and abnormal batch in the 1st heating zone, Figs. 3(b) and 3(c) are the monitoring plots of the 2nd heating zone and soaking zone. The statistics of the batches exceed corresponding control limits which means there are abnormal running states. From Fig. 3, we can observe the monitoring performance of every testing batch in three combustion zones, which helps us to sort out the normal, mild, and heavy abnormal batches. For instance, the batch 32 and batch 33 are heavy abnormal batches since their statistics exceed control limits in two or three subfigures. Similarly, the batch 10 and batch 41 are mild abnormal and normal batches respectively because the overruns occur only in one or no subfigure.
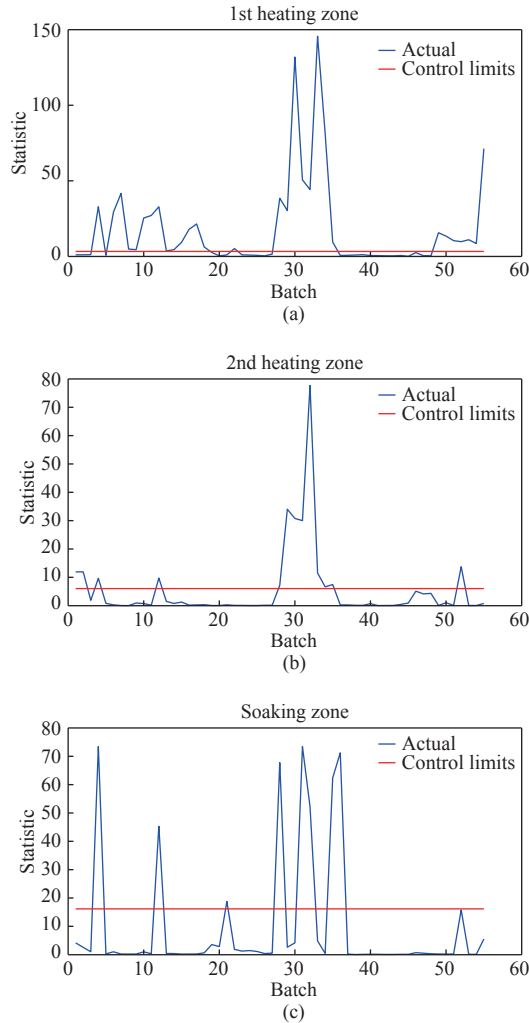
Fig. 4 more clearly and vividly demonstrates the monitoring results of all testing batches from the 3D view, in which all batches of the three zones are obviously displayed as bars, and their statistic values are marked by the color level of the bar sidewall. If a batch of statistics in a certain combustion zone exceeds the corresponding control limit plane, then it indicates that abnormal behavior has occurred. For example, the batch 30 and batch 12 are heavy abnormal batches because in three or two combustion zones there are behaviors that exceed the control limit plane. Since only in one or no combustion zone do their statistics exceed the control limit plane, batch 19 and batch 1 are mild abnormal and normal batch.
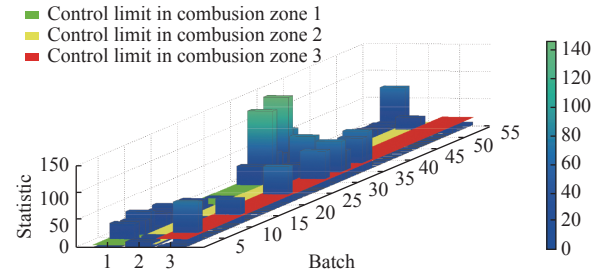

Fig. 4. The whole 3D monitoring charts for all batches.

In order to further demonstrate the superiority of the proposed MDFA-MKECA, it is necessary to do a comparison among MDFA-MKECA, DTW-MKECA, and MDFA-MKPCA. The purpose of comparison between MDFA-MKECA and DTW-MKECA is to show the data pre-processing based on MDFA has more merits than DTW. And the comparison between MDFA-MKECA and MDFA-MKPCA proves that MKECA has a better performance than MKPCA. The results using a percentage of the detection rate (DR) and the abnormality false alarm rate (FAR) as well as the abnormality detection omission rate (DOR) are displayed in Table II. The detection rate includes the normal batches detection rate (NDR), mild abnormality detection rate (MADR), and heavy abnormality detection rate (HADR). DR, FAR, and DOR are defined as follows:

$$R_d = \frac{B_{\text{det}}}{B_{\text{real}}} \times 100\% \qquad (26)$$

where R_d is the detection rate, B_{det} is the number of batches detected (normal, mild abnormality, heavy abnormality), and B_{real} is the numbers of batches that match real production conditions.

$$R_f = \frac{B_{\text{detf}}}{B_{\text{all}}} \times 100\% \qquad (27)$$

where R_f is the abnormality false alarm rate, B_{detf} is the numbers of batches that is detected as abnormal but actually is normal, B_{all} is the total numbers of all testing batches.

$$R_{\text{do}} = \frac{B_{\text{deta}}}{B_{\text{all}}} \times 100\% \qquad (28)$$

where R_{do} is the abnormality omission-detection rate, B_{deta} is the numbers of batches that is detected as normal but actually

Fig. 3. The monitoring charts of all batches in different combustion zones.

TABLE II
THE MONITORING RESULTS AMONG DIFFERENT METHODS

Method	DR			FAR	DOR
	NDR	MADR	HADR		
MDFA-MKECA	86.36%	90.91%	100%	5.45%	3.64%
DTW-MKECA	54.55%	45.45%	100%	34.5%	5.45%
MDFA-MKPCA	77.27%	90.91%	90.91%	9.09%	3.64%

is abnormality, B_{all} is the numbers of all testing batches. Higher DR, lower FAR, and DOR indicate the better process monitoring performance. Table II shows that MDFA-MKECA has a higher DR, and a smaller FAR and DOR than other methods. For the same amount of testing data, the less processing time is used for MDFA-MKECA than DTW-MKECA since the size of data is reduced dramatically by data features extraction.

Except for great monitoring performance, the proposed method also has a merit where the total amount of modeling data is minimal. The MDFA-KECA monitors the process variation in a timely and effective fashion, only using 15 batches modeled. This indicates that the system running state can be represented by a small amount of modeling data and that the proposed MDFA-KECA method has an edge on describing the inner mechanisms, process knowledge and essential information contained in the original data.

VI. CONCLUSION

In this paper, a novel process monitoring method based on mixed data features analysis and multi-way kernel entropy component analysis is applied in modeling a reheating furnace process. The superiority of MDFA-MKECA is that it can detect an abnormal running state effectively and handle complex characteristics among different variables. It proves that mixed data features extraction is an efficient method for capturing original information and reducing the amount of calculation. At the same time, MKECA has a great monitoring performance for batch processes. Extensive simulation results with a reheating furnace process reveal that the proposed method is an appropriate approach to process monitoring.

REFERENCES

[1] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.

[2] Y. Gao, X. Wang, Z. Wang, and L. Zhao, "External analysis and moving window loop based monitoring for a class of multi-mode timevarying process," *IEEE/CAA J. Autom. Sinica*, pp. 1–10, 2017.

[3] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[4] K. X. Peng, M. A. Liang, and Z. Kai, "Review of quality-related fault detection and diagnosis techniques for complex industrial processes," *Acta Automatica Sinica*, vol. 43, no. 3, pp. 349–365, 2017.

[5] N. A. A. Majid, M. P. Taylor, J. J. J. Chen, M. A. Stam, A. Mulder, and B. R. Young, "Aluminium process fault detection by multiway principal component analysis," *Control Engineering Practice*, vol. 19, no. 4, pp. 367–379, 2011.

[6] S. Stubbs, J. Zhang, and J. Morris, "Multiway interval partial least squares for batch process performance monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 35, pp. 12399–12407, 2013.

[7] G. Hui and L. Hongguang, "On-line batch process monitoring with improved multi-way independent component analysis," *Chinese J. Chemical Engineering*, vol. 21, no. 3, pp. 263–270, 2013.

[8] J.-M. Lee, C. Yoo, and I.-B. Lee, "Fault detection of batch processes using multiway kernel principal component analysis," *Computers & chemical engineering*, vol. 28, no. 9, pp. 1837–1847, 2004.

[9] J. Mori and J. Yu, "Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-gaussian latent subspace projection approach," *J. Process Control*, vol. 24, no. 1, pp. 57–71, 2014.

[10] L. Di, Z. Xiong, and X. Yang, "Nonlinear process modeling and optimization based on multiway kernel partial least squares model," in *Proc. IEEE Winter Simulation Conf.*, 2008, pp. 1645–1651.

[11] P. Chang, P. Wang, and X. Gao, "Microbial pharmaceutical process monitoring based on projection kernel entropy component," *Information and Control*, vol. 43, no. 4, pp. 490–494, 2014.

[12] T. Kourti, "Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions," *J. Chemometrics*, vol. 17, no. 1, pp. 93–109, 2003.

[13] P. Nomikos and J. F. MacGregor, "Multivariate spc charts for monitoring batch processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, 1995.

[14] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[15] K. Zhang, K. Peng, and J. Dong, "A common and individual feature extraction-based multimode process monitoring method with application to the finishing mill process," *IEEE Trans. Industrial Informatics*, vol. 14, no. 11, pp. 4841–4850, 2018.

[16] K. Zhang, H. Hao, Z. Chen, S. X. Ding, and K. Peng, "A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches," *J. Process Control*, vol. 33, pp. 112–126, 2015.

[17] Y. J. Jang and S. W. Kim, "An estimation of a billet temperature during reheating furnace operation," *Int. J. Control, Automation, and Systems*, vol. 5, no. 1, pp. 43–50, 2007.

[18] C. Miao, H.-G. Chen, X. Li, and W. Bin, "Total heat exchange factor based on non-gray radiation properties of gas in reheating furnace," *J. Iron and Steel Research, International*, vol. 16, no. 3, pp. 27–63, 2009.

[19] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AIChE J.*, vol. 57, no. 1, pp. 107–121, 2011.

[20] Z. Jj, "Thermodynamic entropy and information entropy," *J. Wuhan University of Science and Technology (Natural Science Edition)*, vol. 1988, no. 4, pp. 21–23, 1988.

[21] D. Aguado, A. Ferrer, J. Ferrer, and A. Seco, "Multivariate spc of a sequencing batch reactor for wastewater treatment," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 1, pp. 82–93, 2007.

[22] S. Wold, N. Kettaneh, H. Fridén, and A. Holmberg, "Modelling and diagnostics of batch processes and analogous kinetic experiments," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1–2, pp. 331–340, 1998.

[23] Y. Yinghua, L. Huaqing, L. Chenlong, Q. Shukai, and C. Xiaobo, "Kernel entropy component analysis based process monitoring method with process subsystem division," in *Proc. 27th IEEE Conf. Control and Decision*, 2015, pp. 2684–2688.

Yinghua Yang received the B.S. degree from Northeastern University in 1991 and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University in 1994 and 2002, respectively. He is an Associate Professor in the College of Information Science and Engineering, Northeastern University. His research interests include process monitoring, fault diagnosis, and data driven modeling methods and their application on smart manufacturing.

Xiaozhi Liu received the B.S. degree from the Shenyang University of Technology in 1990, and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University in 1995 and 2005, respectively. She is an Associate Professor in the College of Information Science and Engineering, Northeastern University. Her research interests include signal processing theory, process monitoring, and fault diagnosis.

Xiang Shi received the bachelor degree in 2015 from Qingdao University. He is currently a master degree candidate in Northeastern University. His main research interests include process monitoring and fault diagnosis based on statistical analysis theory.

Hongru Li received the B.S. degree from Northeastern University in 1991 and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University in 1994 and 2002, respectively. He is a Professor in the College of Information Science and Engineering, Northeastern University. His research interests include modeling, control and optimization methods, process monitoring, and pre diagnosis and heath management of complex systems.