

Multimode Process Monitoring and Fault Diagnosis Based on Tensor Decomposition

Shanshan Zhao* Kai Zhang^{*,**} Kaixiang Peng^{*,**}
Chuanfang Zhang* Xu Yang*

* *Key Laboratory of Knowledge Automation for Industrial Processes of
Ministry of Education, School of Automation and Electrical
Engineering, University of Science and Technology of Beijing, Beijing,
100083, P. R. China (email:18813059729@163.com)*

** *Institute of Artificial Intelligence, University of Science and
Technology Beijing, Beijing 100083, P. R. China*

Abstract: Nowadays, many industrial processes generate large amounts of multimode data, which generally have a natural tensor structure, causing some faults invisible with traditional process monitoring (PM) and fault diagnosis (FD) methods. Tensor decomposition (TD) is a more practical approach for its effectiveness in solving high dimensionality problems as well as indicating the links between different modes. This paper proposes a common and individual feature extraction method based on TD, which identifies and separates the common and individual features from multimode data. The newly proposed approach is applied to a typical multimode hot strip mill process (HSMP), where common and individual feature for all steel products are existing. The final results indicate that the proposed approach can accurately detect and identify different faults in the HSMP.

Keywords: Multimode, tensor decomposition, process monitoring, fault diagnosis, common feature, individual feature.

1. INTRODUCTION

Modern industrial processes have the characteristics of changeable conditions, numerous procedures and complex levels. If a fault cannot be detected and eliminated in time, it may spread during the process, thus affecting the product quality, even causing catastrophic consequences. In order to ensure the safety and stability of production process, process monitoring (PM) and fault detection (FD) have received considerable attention in the last few decades (Venkatasubramanian et al., 2003; Gao et al., 2015).

In modern industrial processes such as the hot strip mill process (HSMP), product quality is determined by a series of working units, including many stands connected via a network, each stand also including different subsystems. The quality of the exit steel can ultimately be influenced by the operating conditions of each stand (Peng et al., 2013, 2014). The data-driven PM method, which is based on the collected process data and uses various data analysis and processing approaches to mine the hidden information and obtain the process state so as to guide the production operation, is more versatile than other PM methods for the ability to deal with the nonlinearity and time-varying characteristics of the process, and has been widely studied and applied in industrial processes (Ge et al., 2013; Shen et al., 2014).

¹ This work was supported by the the Fundamental Research Funds for the Central Universities (Grant No. FRF-TP-19-042A2) the National Key R&D Program of China (Grant No. 2017YFB0306403), and the Natural and Science Foundation of China (NSFC) (Grant No. 61703036, 61673053 and 61873024).

Most of the existing PM methods about industrial processes are operated in a single mode. However, due to complex factors such as material changes, market demands, equipment aging, and seasonal changes, there are often multiple modes in actual processes. Multimode PM requires more complex data, which makes traditional multivariate statistical process control methods no longer applicable, and thus different multimode methods have been developed for multimode PM (Ge et al., 2010; Wang et al., 2012; Zhang et al., 2016; Zhou et al., 2016). Common to these methods is that they only capture the individual feature of each mode data or only account for correlated features within multimode data, which limits their practical applications.

It is not hard to find out that such multimode data naturally share some common features as well as exhibit their own individual features. Intuitively, the common information shared by all modes helps to discover connections between modes while the individual information helps to identify each mode, both of their dimensionality are much lower than the original data. For example, in an HSMP, the mode is often switched from one to another, corresponding to processes producing different steel products. The different modes share some common information, such as the trend of rolling forces and the change of the rolling speed. Besides, there also exists some individual information representing the specific property of each type of steel. Hence, if we can capture the common and individual features of multiple modes, and alternatively monitor the two subspaces, detailed fault information could be acquired for further PM and FD.



Fig. 1. The idea of common and individual feature extraction

The rest of this paper is structured in the following. Section 2 formulate the problems and the basic idea of common and individual feature extraction. Section 3 propose the common and individual feature extraction method using TD. Section 4 develops process monitoring and fault diagnosis methods based on common and individual feature. Section 5 examines the applicability of the new approach to an actual HSMP. The final conclusions will be drawn in Section 6.

2. PROBLEM FORMULATION AND BASIC IDEAS

During the processes to generate different steel slabs in an HSMP, the rolling force variables follow the similar trajectory and can be classified into the common information, while the bending force variable that mainly affect the slab surface quality should be specifically designed for each steel. The common features can significantly determine the operation state which should be monitored in particular. We consider a set of matrices formulated by D samples, M process variables and N operating modes of the HSMP: $\mathcal{X} = \{\mathbf{X}_n \in \mathcal{R}^{D \times M} : n = 1, 2, \dots, N\}$. The measurement space \mathbf{X}_n can be partitioned into common and individual subspaces, which can be shown as

$$\mathbf{X}_n = \bar{\mathbf{X}}_n + \check{\mathbf{X}}_n \quad (1)$$

where $\bar{\mathbf{X}}_n$ represents the common subspaces shared by the N modes, and $\check{\mathbf{X}}_n$ represents the individual subspaces of each mode.

Under this framework, we seek to find

$$\begin{aligned} \mathbf{X}_n &\approx \mathbf{A}_n \mathbf{B}_n^T = [\bar{\mathbf{A}}_n \quad \check{\mathbf{A}}_n] \begin{bmatrix} \bar{\mathbf{B}} \\ \check{\mathbf{B}}_n \end{bmatrix} \\ &= \bar{\mathbf{A}}_n \bar{\mathbf{B}}^T + \check{\mathbf{A}}_n \check{\mathbf{B}}_n^T = \bar{\mathbf{X}}_n + \check{\mathbf{X}}_n \end{aligned} \quad (2)$$

where $\bar{\mathbf{B}} \in \mathcal{R}^{D \times C}$ represents the common feature within the modes while $\check{\mathbf{B}}_n \in \mathcal{R}^{D \times (M-C)}$ represents the individual feature of each slices.

In this way, each mode data \mathbf{X}_n is represented through a combination of components from the common subspace $\bar{\mathbf{X}}_n = \bar{\mathbf{A}}_n \bar{\mathbf{B}}^T$ and the individual subspace $\check{\mathbf{X}}_n = \check{\mathbf{A}}_n \check{\mathbf{B}}_n^T$, as illustrated in Fig. 1.

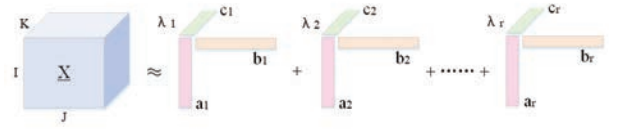


Fig. 2. CPD of a three-order tensor $\underline{\mathbf{X}}$

3. COMMON AND INDIVIDUAL FEATURE EXTRACTION

3.1 Common Feature Extraction Based on the Canonical Polyadic Decomposition (CPD)

A novel approach to calculate the common feature can be achieved by TD, which aims to factorize a data tensor into physically interpretable or meaningful factor matrices and a single core tensor that indicates the links between components in different modes, is more practical in solving high dimensionality problems (Cichocki et al., 2015; Lathauwer, 2008). CPD is the most practical method for TD, which factorizes a tensor into a sum of component rank-one tensors. The CPD of a third-order tensor of rank- R is given by

$$\underline{\mathbf{X}} \approx \sum_{r=1}^R \underline{\mathbf{X}}_r \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\mathbf{A}, \mathbf{B}, \mathbf{C}] \quad (3)$$

If the columns corresponding to the decomposed matrices \mathbf{A} , \mathbf{B} , and \mathbf{C} are regularized, there is a weight vector λ after the decomposition, which is shown in Fig. 2. The formula above is converted as

$$\begin{aligned} \underline{\mathbf{X}} &\approx \sum_{r=1}^R \underline{\mathbf{X}}_r \approx \sum_{r=1}^R \lambda_r \cdot \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \\ &\approx \underline{\Lambda} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = [\underline{\Lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}] \end{aligned} \quad (4)$$

where $\underline{\Lambda}$ is a superdiagonal core tensor that guarantees “one to one relation” for the factor vectors \mathbf{a}_r , \mathbf{b}_r and \mathbf{c}_r , while \mathbf{A} , \mathbf{B} and \mathbf{C} are factor matrices which are composed of the corresponding factor vectors, e.g. $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R]$ and likewise for \mathbf{B} and \mathbf{C} . The sub-matrix $\bar{\mathbf{B}}$ contains the common components shared by all the frontal slices of $\underline{\mathbf{X}}$. And thus, we can define $\bar{\mathbf{B}} = \mathbf{B}$ for further calculation.

The CPD format can be written in matricized form as

$$\begin{aligned} \mathbf{X}_{(1)} &= \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T \\ \mathbf{X}_{(2)} &= \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T \\ \mathbf{X}_{(3)} &= \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T \end{aligned} \quad (5)$$

Alternating Least Squares (ALS) is the most used algorithm for CPD application. The detailed algorithm of a three-order tensor is shown in Table 1 (Kolda et al., 2009).

After calculating the parameters of CPD, the common subspace is defined as

$$\bar{\mathbf{X}}_n = \mathbf{A} \mathbf{C}^{(n)} \mathbf{B}^T \quad (6)$$

where $\mathbf{C}^{(n)} = \text{diag}(\mathbf{c}_n)$, for $n = 1, 2, \dots, N$, and $\bar{\mathbf{B}} = \mathbf{B}$.

Table 1. CPD-ALS algorithm

1: Given: Tensor $\underline{\mathbf{X}}$, rank R
2: Initialization: $\mathbf{A} \in \mathcal{R}^{N \times R}$, $\mathbf{B} \in \mathcal{R}^{M \times R}$, $\mathbf{C} \in \mathcal{R}^{K \times R}$
3: Repeat
$\mathbf{A} = \mathbf{X}_{(1)}(\mathbf{C} \circledast \mathbf{B})(\mathbf{B}^T \mathbf{B} * \mathbf{C}^T \mathbf{C})^\dagger$
$\mathbf{B} = \mathbf{X}_{(2)}(\mathbf{C} \circledast \mathbf{A})(\mathbf{A}^T \mathbf{A} * \mathbf{C}^T \mathbf{C})^\dagger$
$\mathbf{C} = \mathbf{X}_{(3)}(\mathbf{B} \circledast \mathbf{A})(\mathbf{A}^T \mathbf{A} * \mathbf{B}^T \mathbf{B})^\dagger$
normalize columns of \mathbf{A} , \mathbf{B} and \mathbf{C} (storing norms as λ)
Until fit ceases to improve or maximum iterations exhausted
4: Output: $\lambda, \mathbf{A}, \mathbf{B}, \dots, \mathbf{C}$

3.2 Individual Feature Extraction Based on Principal Component Analysis (PCA)

After figuring out the common subspaces, PCA can be applied on the rest parts $\mathbf{X}_n - \bar{\mathbf{X}}_n = \check{\mathbf{X}}_n$, which is expressed as

$$\check{\mathbf{X}}_n = \check{\mathbf{A}}_n \check{\mathbf{B}}_n^T \quad (7)$$

4. PM AND FD METHODS BASED ON COMMON AND INDIVIDUAL FEATURE

4.1 PM method

Given the new measurements from n -th mode \mathbf{x}_{new} , it is first normalized, then we replace the train samples of the n -th mode with this new observation while keeping the train samples of the other modes fixed. And then, we unfold the new tensor into a matrix from the sample direction to gain the mode-1 factorization $\mathbf{x}_{(1)_{new}}$. The common score is obtained as:

$$\bar{\mathbf{A}}_{new} = \mathbf{x}_{(1)_{new}}[(\mathbf{C} \circledast \mathbf{B})^T]^\dagger \quad (8)$$

And then, the common part can be calculated as:

$$\bar{\mathbf{x}}_{new} = \bar{\mathbf{A}}_{new} \mathbf{C}^{(n)} \mathbf{B}^T \quad (9)$$

After removing the common part, the left part is acquired as:

$$\check{\mathbf{x}}_{new} = \mathbf{x}_{new} - \bar{\mathbf{x}}_{new} \quad (10)$$

The specific score is calculated as:

$$\check{\mathbf{A}}_{new} = \check{\mathbf{x}}_{new} \check{\mathbf{B}}_n \quad (11)$$

Using $\check{\mathbf{A}}_{new}$, the specific part is reconstructed as:

$$\check{\mathbf{x}}_{new} = \check{\mathbf{A}}_{new} \check{\mathbf{B}}_n^T = \check{\mathbf{x}}_{new} \check{\mathbf{B}}_n \check{\mathbf{B}}_n^T \quad (12)$$

Finally, the remained part is obtained as:

$$\mathbf{e}_{new} = \mathbf{x}_{new} - \hat{\mathbf{x}}_{new} \quad (13)$$

The T^2 statistic, which is the projection size of the variables in the principal subspace, reflecting the amount of changes of each variable, can be used to monitor common subspaces. The mathematical expression is

$$T_c^2 = \bar{\mathbf{A}}_{new} \left(\frac{1}{n-1} \bar{\mathbf{A}}_{new}^T \bar{\mathbf{A}}_{new} \right)^\dagger \bar{\mathbf{A}}_{new}^T \quad (14)$$

$$T_i^2 = \check{\mathbf{A}}_{new} \left(\frac{1}{n-1} \check{\mathbf{A}}_{new}^T \check{\mathbf{A}}_{new} \right)^\dagger \check{\mathbf{A}}_{new}^T \quad (15)$$

The control limit of T^2 is expressed as

$$T_\alpha^2 = \frac{A(n^2 - 1)}{n(n - A)} F_{A, n-A, \alpha} \quad (16)$$

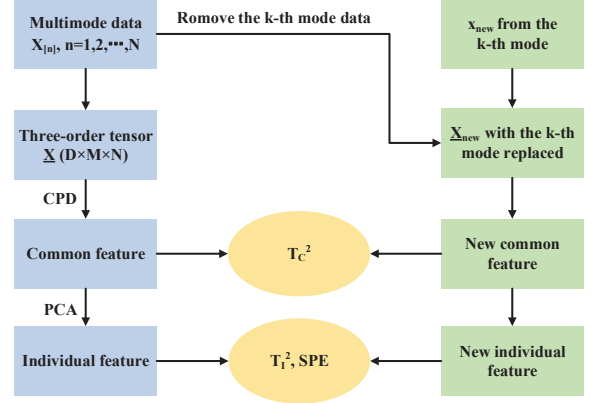


Fig. 3. A schematic illustration of the proposed monitoring approach

where $F_{A, n-A, \alpha}$ represents the critical value of the F distribution with A and $n - A$ degrees of freedom and a confidence level of α .

Since the individual subspaces use PCA, it can be monitored by the T^2 and squared prediction error (SPE) statistic, which start from the linear relationship between variables to detect whether the system is faulty or not (Peng et al., 2011).

The SPE statistic reflects the changes in the projection of the data sample vector in the residual subspace, that is

$$SPE = \|\mathbf{e}_{new}\|^2 \quad (17)$$

The control limit of SPE is expressed as

$$\delta_\alpha^2 = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (18)$$

where $\theta_i = \sum_{j=A+1}^m \lambda_j^i$ ($i = 1, 2, 3$), $h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$, λ_j represents the i -th eigenvalue of the data sample covariance matrix, c_α represents the threshold of the standard normal distribution under confidence α .

If T_c^2 is larger than the threshold, a structural fault causing changes to the common feature has occurred. If only T_i^2 or SPE are larger than their thresholds, a fault that affects only the specific characteristics of this mode occurred.

The schematic illustration of the proposed method is shown in Fig. 3.

4.2 FD method

However, when the SPE and T^2 statistics exceed their control limits, they can only determine that the process has failed. It is impossible to find out the fault happened in which part of the HSMP from the SPE and T^2 figures. The contribution plot is a simple and practical FD method, which reflects the influence of each variable on the stability of the system statistical model. The core idea of the method is that the high contribution value of the process variable can be considered as the cause of the fault. Combining these analytical results with the knowledge of the HSMP will make it easier to identify the cause of the

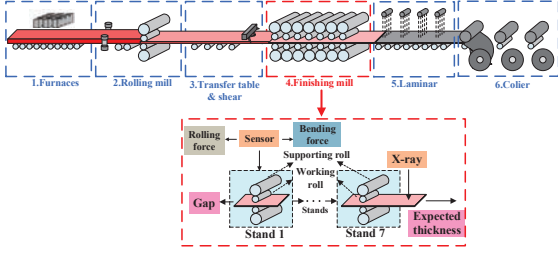


Fig. 4. The layout of the hot rolling mill process

failure. The T^2 contribution value is represented by the contribution of each variable to the score:

$$c_i^{T^2} = (\xi_i^T (\bar{\mathbf{B}} \mathbf{\Lambda}_c^{-1} \bar{\mathbf{B}}^T)^{1/2} \bar{\mathbf{x}}_{new}) \quad (19)$$

$$c_i^{T_i^2} = (\xi_i^T (\check{\mathbf{B}} \mathbf{\Lambda}_i^{-1} \check{\mathbf{B}}^T)^{1/2} \check{\mathbf{x}}_{new}) \quad (20)$$

where ξ_i is the i -th column of the identity matrix, .

5. APPLICATION TO THE HSMP

5.1 Description of the HSMP

Modern HSMP is an automated production line running in the order form and the production line is mainly composed of furnaces, rolling mill, transfer table and shear, finishing mill, laminar cooling and coiler. The steel slab is first heated up to about 1200°C in the reheating furnace. The thickness and the length of hot steel slab are shaped roughly to the expectations during the roughing mill process (RMP). The finishing mill process (FMP) gives further precise gauge reduction towards the preset width and thickness. Then the extremely hot strip steels are cooled to the desired temperature by laminar-cooling equipment and are finally coiled by the coiler for convenient loading. An HSMP consists of a sequence of mill stands, which work collectively to mill the raw steel slab to the required specifications. At each stand, two types of rollers are equipped. Two kinds of forces are applied to rollers to ensure that the steel thickness can be reduced as expected, and the steel crown that quantifies the thickness across the width of the strip can reach an acceptable level. Four rollers are rotated by driven motors so that the steel slab can smoothly move forward. Meanwhile, the steel temperature at the exit of an HSMP should strictly fall into a certain range to ensure the required microstructure property inside the steel. The layout of an hot rolling mill process is shown in Fig. 4.

The FMP plays an important role in HSMP, by which a hot strip steel coming from the rough mill process can be precisely milled to meet the technical requirement. There are seven stands in FMP, of which each stand has two supporting rolls in two sides and two working rolls in the middle. Before a strip steel enters a stand, a proper initial rolling force is given to the upper supporting roll by the hydraulic system, which makes the gap between two working rolls less than the thickness of the coming steel. An initial speed is set on the upper working roll via a driving motor. When the head of a strip steel reaches two working rolls, the actual gap will be enlarged due to the rebounding phenomenon. An electromechanical system is placed to rotate the rolls so that the strip

steel can be smoothly moved forward. Meanwhile, the rolling and bending forces can be measured in real time by piezomagnetic and strain gauge sensors. In practice, the gap between the two working rolls can be a proxy variable that can be measured by the height difference between the two supporting rolls, and the exit thickness can be measured by an X-ray device with time delay due to it being placed far from the exit of the last stand (Ma et al., 2018; Zhang et al., 2018).

5.2 Data preparation

The data come from the database of a real steel plant. We consider three operating modes, one of which producing 2.70-mm strip steels and the other two producing 3.95-mm strip steels. We truncate a period covering 35s of normal operations, with a sampling interval of 0.1s. A total of 3500 samples of 20 process variables of roll gap, rolling force and bending force of seven stands (1st stand without bending rolls) were collected as $X_1 \in \mathcal{R}^{3500 \times 20}$, $X_2 \in \mathcal{R}^{3500 \times 20}$, and $X_3 \in \mathcal{R}^{3500 \times 20}$.

5.3 Fault scenarios

Three typical fault scenarios are taken into consideration. The first fault is a structural fault, which arises due to the fault of the hydraulic system in the fourth stand. This fault will immediately propagate to the downstream stands and finally affect the exit thickness, which has both impacted the common and individual features. The second fault is a non-structural fault, which occurs in the bending force sensor at the fifth stand. The influence of this fault can be compensated by enlarging the bending force in sixth and seventh stands, thus will not affect the exit thickness and will not affect the common feature. The third fault is also a structural fault, arising by blocking of spray water valve located between second and third stands. It increases the surrounding temperature, then the downstream stands are affected due to the feedforward controller and can both affect the common and individual features.

5.4 Simulation results

The data of the three modes can be organized as a third-order tensor $\mathbf{X} \in \mathcal{R}^{3500 \times 20 \times 3}$. CPD is applied to \mathbf{X} . However, determining the rank- R of a tensor \mathbf{X} is often a difficult problem. There is no straightforward algorithm to determine the rank of a specific given tensor. In this paper, a multilinear singular values-based method is used to help choosing R , by which we can plot an L-curve which represents the balance between the relative error of the CPD $\|\mathbf{X} - \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r\| / \|\mathbf{X}\|$ and the number of rank-one terms R . The algorithm computes the CPD of the given tensor for various values of \mathbf{X} , starting at the smallest integer for which the lower bound on the relative error is smaller than the MAX option. The number of rank-one terms is increased until the relative error of the approximation is less than MIN option. It is recommended to set MAX higher than the expected noise level and to set MIN at a value near the noise level and smaller than MAX , as is shown in Fig. 5, the dimension of the common subspace R is set as 2.

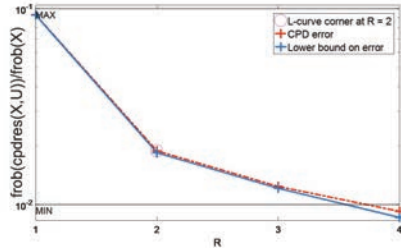


Fig. 5. The *rankest* tool for choosing the number of R in a CPD

Table 2. FDR of fault case 1 when R varies

R	1	2	3	4	5
T_c^2	0.864	1.000	0.989	0.891	0.883
T_i^2	0.934	1.000	1.000	1.000	0.991
SPE	1.000	1.000	1.000	1.000	1.000

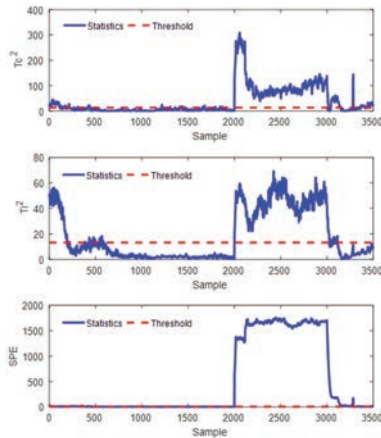


Fig. 6. Monitoring results of fault 1

According to the result, the best R value is 2. Nevertheless, we also need to verify this set. We make R=1, 2, 3, 4, 5 respectively, and then carry out CPD on the three-order tensor \mathbf{X} , and extract common and individual features for PM. Take fault case 1 for example, the fault detection rate (FDR) when R varies is shown in Table 2, which confirms the previous inference.

The monitoring results of the three fault scenarios are shown in Figs. 6-8. It can be seen that fault 1 occurs at the 2000th sample, and all three statistics have detected this fault. Fault 2 arises at the 1000th sample, and only T_i^2 and SPE statistics have detected this fault while T_c^2 does not have any reaction to this fault. Fault 3 occurs at the 1000th sample, and all three statistics have detected the fault, which are consistent with the former descriptions.

The FDR and fault alarm rate (FAR) can be used to compare the PM results between different methods. The compared fault detection rate of the proposed TD-PCA-based method and multimode PCA is shown in Table 2 and Table 3. From the compared fault detection rate, we can see that the PM of common and individual subspaces has better results than traditional PCA methods.

The T^2 contribution plots of the three fault cases as shown in Figs. 9-11. It can be seen that the sources of fault scenario 1 roughly locate at the rolling force and gap in

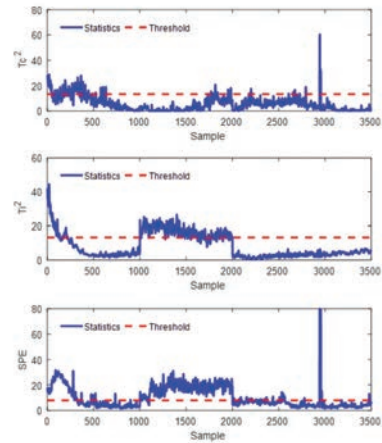


Fig. 7. Monitoring results of fault 2

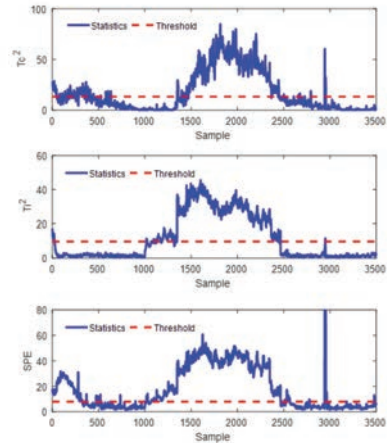


Fig. 8. Monitoring results of fault 3

Table 3. FDR by TD-PCA-based method and multimode PCA

Methods	Monitoring Statistics	Fault 1	Fault 2	Fault 3
TD-PCA	T_c^2	1.000	0.015	0.824
	T_i^2	1.000	0.915	0.955
	SPE	1.000	0.975	0.949
Multimode PCA	T^2	1.000	0.129	0.521
	SPE	1.000	1.000	0.924

Table 4. FAR by TD-PCA-based method and multimode PCA

Methods	Monitoring Statistics	Fault 1	Fault 2	Fault 3
TD-PCA	T_c^2	0.102	0.094	0.116
	T_i^2	0.128	0.097	0.012
	SPE	0.017	0.162	0.109
Multimode PCA	T^2	0.124	0.113	0.131
	SPE	0.112	0.136	0.127

stand 6 in the common subspace and rolling force in stand 4 and gap in stand 4 in the individual subspace. The main sources of fault scenario 2 are the bending force in stand 5 and stand 6. And the fault variables in fault scenario 3 are the rolling force in stand 5 and gap in stand 3-6 in the common subspace and rolling force in stand 3 in the individual subspace.

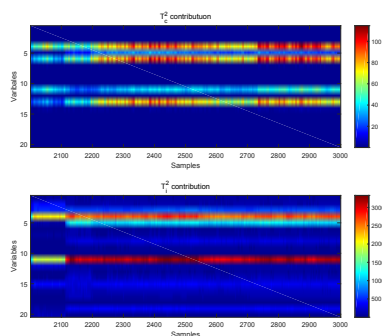


Fig. 9. Contribution plot of fault 1

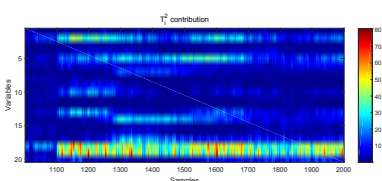


Fig. 10. Contribution plot of fault 2

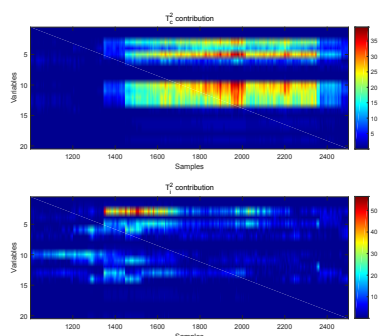


Fig. 11. Contribution plot of fault 3

6. CONCLUSION

In this paper, a new common and individual feature extraction PM method based on TD was proposed for multimode data. The common feature of different mode data can be extracted by CPD and the individual feature of each mode data can also be calculated. The new framework was realized using T^2 and SPE statistics for PM and using contribution plot for fault diagnosis. The proposed methods have been applied to a multimode HSMP, where three typical fault cases were taken into account. Using the proposed approach, faults that affect the common and individual features can be accurately detected than traditional methods.

REFERENCES

Venkatasubramanian, Venkat and Rengaswamy, Raghunathan and Yin, Kewen and Kavuri, Surya N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers and Chemical Engineering*, 27(3):293–311, 2013.

Gao, Zhiwei and Cecati, Carlo and Ding, Steven X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part I: Fault Diagnosis With Model-Based and Signal-

Based Approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3757–3767, 2015.

Peng, Kaixiang and Kai, Zhang and Gang, Li and Zhou, Donghua. Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process. *Control Engineering Practice*, 21(4):360–369, 2013.

Peng, Kaixiang and Hao, Zhong and Liang, Zhao and Kai, Xue and Ji, Yidao. Strip shape modeling and its setup strategy in hot strip mill process. *International Journal of Advanced Manufacturing Technology*, 72(5-8):589–605, 2014.

Ge, Zhiqiang and Song, Zhihuan and Gao, Furong. Review of Recent Research on Data-Based Process Monitoring. *Industrial and Engineering Chemistry Research*, 52(10):3543–3562, 2013.

Shen, Yin and Ding, S. X. and Xie, Xiaochen and Hao, Luo. A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE Transactions on Industrial Electronics*, 61(11):6418–6428, 2014.

Ge, Zhiqiang and Song, Zhihuan. Multimode process monitoring based on Bayesian method. *Journal of Chemometrics*, 23(12):636–650, 2010.

Wang, Fuli and Shuai, Tan and Peng, Jun and Chang, Yuqing. Process monitoring based on mode identification for multi-mode process with transitions. *Chemometrics and Intelligent Laboratory Systems*, 110(1):144–155, 2012.

Zhang, Yingwei and Fan, Yunpeng and Nan, Yang. Fault diagnosis of multimode processes based on similarities. *IEEE Transactions on Industrial Electronics*, 63(4):2606–2614, 2016.

Cichocki, Andrzej and Mandic, Danilo and De Lathauwer, Lieven. Tensor Decompositions for Signal Processing Applications From Two-way to Multiway Component Analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, 2015.

De Lathauwer, Lieven. Decompositions of a Higher-Order Tensor in Block Terms Part III: Alternating least squares algorithms. *Siam Journal on Matrix Analysis and Applications*, 30(3):1067–1083, 2008.

Zhou, G. and Cichocki, A and Zhang, Y. and Mandic, D. P. Group Component Analysis for Multiblock Data: Common and Individual Feature Extraction. *IEEE Trans Neural Netw Learn Syst*, 27(11):2426–2439, 2016.

Kolda, Tamara G. and Bader, Brett W. Tensor Decompositions and Applications. *SIAM Review*, 51(3):455–500, 2009.

Peng, Kai Xiang and Zhou, Dong Hua and Li, Na. Quality-Related Monitoring and Control in Hot Strip Mill Process. *Control Engineering of China*, 18(4):650–654, 2011.

Ma, Liang and Dong, Jie and Peng, Kaixiang and Zhang, Chuanfang. Hierarchical Monitoring and Root Cause Diagnosis Framework for Key Performance Indicator-Related Multiple Faults in Process Industries. *IEEE Transactions on Industrial Informatics*, 15(4):2091–2100, 2018.

Zhang, Kai and Peng, Kaixiang and Dong, Jie. A Common and Individual Feature Extraction-Based Multimode Process Monitoring Method with Application to the Finishing Mill Process. *IEEE Transactions on Industrial Informatics*, 14(11):4841–4850, 2018.