Research and application of KICA-AROMF based fault diagnosis

Qun-Xiong Zhu, Qian-Qian Meng, Yuan Xu, Yan-Lin He*

Abstract—With the development of the modern industrial system, data-driven fault diagnosis methods have attracted more and more attention. Fault diagnosis of complex industrial processes based on one-dimensional adaptive rank-order morphological filter (AROMF) may miss key information because of excessive dimension reduction of process data. To solve this problem, a method combining the kernel independent component analysis (KICA) with one-dimensional AROMF is proposed. Firstly, KICA is used for nonlinear feature extraction, getting the template signal and the test signal of each pattern. Then, a fault diagnosis method via multi-dimensional signals classification method based on AROMF is presented in this paper. The advantage of the proposed method was confirmed by the simulation of the Tennessee Eastman process.

I. INTRODUCTION

Effective fault diagnosis is becoming more and more important owing to the automatic, intelligent trend of industry systems. With the establishment and constant improvement of the theory and technology, fault diagnosis methods can be divided into three categories: model based methods, knowledge based methods and data based methods [1, 2]. Due to the development of artificial intelligence and computer technology, the process data generated by complex industrial systems can be fully utilized. Therefore, data based methods that don’t require accurate mathematical model have become a popular research field in recent years [3].

In order to get better use of the high dimensional, high correlation and serious nonlinear process data generated by the complex industrial system, many feature extraction techniques have been applied. Such as auto-associative neural network (AANN) [4, 5], principle component analysis (PCA) [6], independent component analysis (ICA) [7], etc. ICA is the extension of the PCA and factor analysis [8]. ICA can be used to extract the feature information of non-Gaussian distribution samples. ICA can also reduce the correlation between variables [9]. However, if the data is Gauss distribution or near Gauss distribution, feature extraction based on ICA will lead to increases of false positive rate and false negative rate [10]. Then, for improving the above problem, a new algorithm, kernel-ICA (KICA) based on the kernel function theory is proposed by F R Bach et al. [11]. References [12, 13] indicated that KICA can achieve excellent performance for data feature extraction of complex system fault diagnosis and on-line monitoring. Taking into account the characteristics of the process data of industrial system, this paper used KICA algorithm to extract feature of data.

Noisy process data collected under different fault patterns are certainly similar. The rank-order morphological filter (ROMF) based on mathematical morphology is a kind of signal processing method with good filtering effect. The filtering characteristic is directly related to the structuring elements and the percentile [14]. In paper [15] the improved order morphological filter algorithm is applied to the signal de-noising in the mechanical field. An adaptive rank-order morphological filter (AROMF) is proposed by Han Li etc. [16]. One-dimensional AROMF can make the structuring elements and the percentile of each sampling point adjusted adaptively. As a kind of supervised transformation technology, one-dimensional AROMF can reduce the difference between the supervised signal and the test signal in the similar state.

Process data carries comprehensive system information, if only one-dimensional signal is extracted, some key information may be missed. Considering this problem, and in order to avoid the problem of missing different contribution of each dimensional signal for the system in different state, a KICA-AROMF based fault diagnosis was proposed in this paper.

II. PRINCIPLE OF KERNEL INDEPENDENT COMPONENT ANALYSIS

The main idea of KICA is mapping the original data into high dimensional feature space \( \Phi \) via the nonlinear function \( \phi(\cdot) \), so that ICA can deal with these data.

There is the original process data \( \textbf{X}=[x_1,x_2,\cdots, x_N] \), \( N \) is the number of variable. \( x_i \) is mapped to \( \phi(x_i) \) via the mapping function \( \phi(\cdot) \), the high dimensional space is represented as \( \Phi=[\phi(x_{i1}), \phi(x_{i2}), \cdots, \phi(x_{iN})] \). Because of the difficulty of determining the mapping function, the kernel function \( k(x_i,x_j) \) is introduced to compute inner products in the feature space:

\[
K_{ij} = \phi(x_{i1})^T \phi(x_{j1})
\]

Radial basis kernel function, polynomial kernel function, S kernel function, etc. are common kernel functions. This paper chooses the radial basis kernel function:

*Research supported by National Natural Science Foundation of China. Qun-Xiong Zhu, was with College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China and the Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China (e-mail: zhuqx@mail.buct.edu.cn).

Qian-Qian Meng, was with College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China and the Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China. Now she is pursuing the master degree (e-mail: 2015200723@mail.buct.edu.cn).

Yuan Xu, was with College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China and the Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China (e-mail: yxyuan@mail.buct.edu.cn).

Yan-Lin He, was with College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China and the Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China (corresponding author to provide phone: +86-10-64426960; fax: +86-10-64437805; e-mail: heyl@mail.buct.edu.cn).
It is necessary to center and whitening of the matrix $K$ before the independent component extraction:

$$I_N = K - I_N K - K I_N + I_N K I_N$$  \hspace{1cm} (3)

$I_N$ is a constant matrix. In this constant matrix all elements are $I/N$. Obtain the first $d$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and corresponding eigenvectors $\alpha_1, \alpha_2, \ldots, \alpha_d$ of $\tilde{K}$ via Eigen-decomposition. Then whitening data is defined as follows:

$$Z = Q\psi\varphi(x) = \sqrt{N}\Lambda^{-1}H^T\tilde{K}x.$$  \hspace{1cm} (4)

In Eq. (4), $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$, $H = [\alpha_1, \alpha_2, \ldots, \alpha_d]$, $Q^\psi = \sqrt{N}\Phi\Lambda^{-1}$ is the whitening matrix, $\tilde{K}x$ is the result of the centralizer of $K_x$, $K_x = \Phi^T\psi(x)$.

Finally, a fast-ICA algorithm is used to extract nonlinear independent components after obtaining the whitening matrix $Z$ in the feature space.

### III. ONE-DIMENSIONAL ADAPTIVE RANK-ORDER MORPHOLOGICAL FILTER (AROMF)

Rank-order morphological filtering (ROMF) based on data sorting, is an extension of the Minkowski structure and the median filter [16]. The main principle of ROMF is shown as follows: after sorting a small segment of signal, select an optimal sample point to represent the entire segment of the signal. ROMF can avoid the limitation of only choosing the maximum or minimum value of the selected signal segment.

Rank-order morphological filtering is defined as: set $f(t)$ is an one dimensional discrete signal, $B = \{j_1, j_2, \ldots, j_{N_B}\}$ is the structural element, $\mu(B) (0 < \mu(B) = N_B < +\infty)$ is the number of element in $B$. Sort values of $f(B)$ defined in $B$ in ascending order, $f(t_{i1}) \leq f(t_{i2}) \leq \cdots \leq f(t_{iN_B})$.

Denote the $Oth$ ordered value of $f(t)$ on $B$ as Eq. (5):

$$\text{ord}(o, f(B) | B) = f(t_{i_o}) \in \{1, 2, \ldots, N_B\}$$  \hspace{1cm} (5)

Then one-dimensional ROMF $f(B)(t)$ can be defined as Eq. (6):

$$(\Theta B)(t) = \text{ord}(o, f(B_t) | B'_p) = \text{ord}((N_B - 1)p + 1, f(B'_p)$$  \hspace{1cm} (6)

$B'_t = \{t - j, j \in B\}$, $p = 0, 1/(N_B - 1), \ldots, 1$ is called rank-order morphological filter percent.

By the above definition can be judged, the main effects of the performance of filter are $\mu(B)$ the number of element in $B$, and $p$ the percentile value of rank-order morphological filter. A novel adaptive algorithm named one-dimensional adaptive rank-order morphological filter (AROMF) is proposed in reference [17]. AROMF can realize parameters $B$ and $p$ updating adaptively. The adaptive ability of AROMF for the signal feature selection and the application is stronger. A detailed description of one-dimensional AROMF can be obtained in reference [17].

### IV. KICA-AROMF BASED FAULT DIAGNOSIS METHOD

#### A. Classification method of multi-dimensional signals based on AROMF

Due to the complexity of the system, one IC cannot reconstruct the complete information of the system state unless there are enough ICs. Considering this, a multi-dimensional signals classification method based on one-dimensional adaptive rank-order morphological filter (AROMF) was proposed.

AROMF is a kind of filtering which regards the desired signal as a supervisory signal to filter the test signal. For the same noise test signal $x$ using different supervisory signals $d$ will have different output signals $y$. If an undesired signal is used as a supervisory signal, the recovery effect of noisy signal will be poor [17]. The distance between output signal $y$ and desired signal $d$ is the smallest.

For a certain test signal the output signal $y_i = \{y_i^1, y_i^2, \ldots, y_i^n\}$ obtained under supervision of $d = \{d_1^n, d_2^n, \ldots, d_i^n\}$, where $i = 1, 2, \ldots, M$ is the number of supervisory signal, which means the number of fault patterns.

Define the distance between the output signal $y_i^n$ and the supervisory signal $d_i^n$ as $E_i^n$, calculated as Eq. (7):

$$E_i^n = \sum_{l=1}^{N} \sqrt{(y_{il} - d_{il})^2}$$  \hspace{1cm} (7)

where $L$ is the number of signal sampling points; $d_i^n$ are multi-dimensional template signals for $i$th fault pattern, $d_i^n$ is the component signal on $n$th IC of $d_i^n$; $y_i^n$ is the output signal obtained under supervision of $d_i^n$.

Define the mean distance between $y_i$ and $d_i$ as $E_i$, calculated as Eq. (8):

$$E_i = \frac{1}{N} \sum_{n} E_i^n$$  \hspace{1cm} (8)

where $N$ is the number of ICs in KICA model, $y_i$ are multi-dimensional output signals obtained under supervision of $d_i$. The flow chart for calculating the mean distance of multi-dimensional signal is shown in Fig. 1.

![Figure 1](image-url)
Therefore, the main principle of the proposed multi-dimensional signals classification method is listed as follows: recovering different new signals $y_i$ from the unknown detection signal $x$ by regarding each template signal of known fault patterns as different supervisory signals $d_i$; then, measure the difference between the output signal $y_i$ and the template signal $d_i$ via Eq. (7) and Eq. (8). The flow chart for multi-dimensional signals classification method is shown in Fig. 2.

![Flow chart for multi-dimensional signals classification method](image)

**Figure 2.** The flow chart for multi-dimensional signals classification method

**B. KICA-AROMF based fault diagnosis**

Considering the characteristics of process data from complex industrial system, KICA was chosen to extract feature signals of each fault pattern. The feature signals should be multi-dimensional, so that they can reconstruct the information of each fault pattern completely. The steps of KICA-AROMF based fault diagnosis method are given.

1) **Data preprocessing:** de-noise and normalization of original training data and test data.

2) **Establish KICA model:** The KICA algorithm is used to extract the ICs of the data from all fault patterns, expressed as: $IC_1, IC_2, \ldots, IC_n$.

3) **Obtain the template signal:** the template signals are obtained by projecting the training signal of each single fault pattern onto the KICA model.

4) **Obtain the test signal:** projecting the original non-classified signal onto the KICA model as the test signal.

5) **Determine fault pattern:** using the multi-dimensional signals classification method based on AROMF mentioned above.

**V. APPLICATION IN TE PROCESS**

This section provides the test of our fault diagnosis method in the TE process [18] to verify its superiority and efficiency.

There are 4 ICs in the established KICA model. After projecting the original training signal of each known fault pattern onto this KICA model, many corresponding template signals are obtained. Four of them are shown in Fig. 3. Comparing template signals of each fault pattern with that of the normal state, we can see that the change of template signals on the same IC but in different fault patterns are different. Hence, it is necessary to obtain multi-dimensional characteristic signals.

There are 12 manipulated variables, and 41 observed variables in the TE process. They can be expressed as XMV(1)-XMV(12) and XOV(1)-XOV(41) respectively. Since XMV(5), XMV(9) and XMV(12) are constants. They will be removed from the original signal. The first 7 of the 21 fault patterns in the TE process are selected as the test object shown in Table I. Experiment process of fault diagnosis is according to the given diagnosis procedure in section 3(B).

<table>
<thead>
<tr>
<th>Fault pattern</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDV(0)</td>
<td>normal state</td>
</tr>
<tr>
<td>IDV(1)</td>
<td>A/C feed ratio, B composition constant (stream 4)</td>
</tr>
<tr>
<td>IDV(2)</td>
<td>B composition, A/C ration constant (stream 4)</td>
</tr>
<tr>
<td>IDV(3)</td>
<td>D feed temperature (stream 2)</td>
</tr>
<tr>
<td>IDV(4)</td>
<td>reactor cooling water inlet temperature</td>
</tr>
<tr>
<td>IDV(5)</td>
<td>condenser cooling water inlet temperature</td>
</tr>
<tr>
<td>IDV(6)</td>
<td>A feed loss (stream 1)</td>
</tr>
<tr>
<td>IDV(7)</td>
<td>C header pressure loss-reduced availability (stream 4)</td>
</tr>
</tbody>
</table>

**TABLE I.**  **FIRST 7 FAULT PATTERNS IN THE TE PROCESS**

(a) IDV(0)
After obtaining template signals, the KICA model is used to detect characteristic signals of original signals to be classified as the test signal. Then determine the fault pattern using the multi-dimensional signals classification method based on AROMF. There are the results of AROMF algorithm for IDV(3) with different supervision signals shown in Fig. 4. The results confirm that the gap between the output signal and the supervisory signal is the smallest when the template signal in the same fault pattern is used as the supervisory signal.
Each fault pattern of IDV (1) ~ IDV (7) simulated by 50 times in this paper. Without loss of generality, 25 sets of test data were collected from starting-up 4 hours to add the fault, and the rest were from starting-up 4 hours to add the fault. Correct rates of different fault diagnosis methods are compared in Table II.

<table>
<thead>
<tr>
<th>Fault pattern</th>
<th>KICA-AROMF</th>
<th>One-dimensional AROMF [17]</th>
<th>KICA [19]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDV(1)</td>
<td>100%</td>
<td>100%</td>
<td>99.75%</td>
</tr>
<tr>
<td>IDV(2)</td>
<td>90%</td>
<td>100%</td>
<td>98.12%</td>
</tr>
<tr>
<td>IDV(3)</td>
<td>66%</td>
<td>30%</td>
<td>2.25%</td>
</tr>
<tr>
<td>IDV(4)</td>
<td>76%</td>
<td>30%</td>
<td>95.13%</td>
</tr>
<tr>
<td>IDV(5)</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>IDV(6)</td>
<td>92%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>IDV(7)</td>
<td>100%</td>
<td>100%</td>
<td>98.6%</td>
</tr>
</tbody>
</table>
Table II suggests that the method proposed in this paper can catch system features of IDV(1), IDV(2), IDV(5), IDV(6) and IDV(7), because they have higher correct rate. When IDV(3) and IDV(4) are added, comparing with the normal state, the change of mean and variance of each variable are not obvious. Thus, IDV(3) and IDV(4) are easy to be misdiagnosed. Comparing with the statistical data in reference [17], the correct rates of IDV(3) and IDV(4) are much higher, however, in reference [19] the IDV(3) can hardly be diagnosed. Since the effect of fault time was taking into account in our experiments, the correct rates of IDV(2) and IDV(6) are affected to a certain extent. Overall, the average correct rate of KICA-AROMF based, one-dimensional AROMF, and KICA based fault diagnosis method is 89.14%, 80%, 84.84%, respectively. Thus, the superiority and efficiency of proposed method are confirmed.

VI. CONCLUSION

A classification method of multi-dimensional signals based on AROMF is proposed in this paper. In different patterns, the contribution of component signals on each IC is different. Hence, Multi-dimensional characteristic signals of complex system carry more comprehensive information. The above factor is taken into account in the proposed classification method. Therefore, the fault pattern that system variables don’t change obviously can also be diagnosed. And the conclusion is proved by simulation experiments of TE process.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China, China Grant no. 61473026 and 61573051.

REFERENCES