



Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Balanced multiple weighted linear discriminant analysis and its application to visual process monitoring

Weipeng Lu, Xuefeng Yan*

Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China



ARTICLE INFO

Article history:

Received 22 July 2020

Received in revised form 15 October 2020

Accepted 28 October 2020

Available online 27 November 2020

Keywords:

Linear discriminant analysis

Process monitoring

Self-organizing map

Feature extraction

Continuous stirred tank reactor process

ABSTRACT

Visual process monitoring is important in complex chemical processes. To address the high state separation of industrial data, we propose a new criterion for feature extraction called balanced multiple weighted linear discriminant analysis (BMW LDA). Then, we combine BMW LDA with self-organizing map (SOM) for visual monitoring of industrial operation processes. BMW LDA can extract the discriminative feature vectors from the original industrial data and maximally separate industrial operation states in the space spanned by these discriminative feature vectors. When the discriminative feature vectors are used as the input to SOM, the training result of SOM can differentiate industrial operation states clearly. This function improves the performance of visual monitoring. Continuous stirred tank reactor is used to verify that the class separation performance of BMW LDA is more effective than that of traditional linear discriminant analysis, approximate pairwise accuracy criterion, max–min distance analysis, maximum margin criterion, and local Fisher discriminant analysis. In addition, the method that combines BMW LDA with SOM can effectively perform visual process monitoring in real time.

© 2020 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights reserved.

1. Introduction

Process monitoring is of great significance for ensuring process safety and for improving product quality. Multivariate statistical methods, such as principal component analysis (PCA) [1], partial least squares [2], and independent component analysis [3], have been widely adopted in process monitoring. In addition, the widely used deep neural networks, such as stacked auto-encoder [4,5], deep belief network [6], convolutional neural network [7], and long short-term memory [8], have also been introduced for process monitoring. However, these methods do not provide the visualization of the industrial process, so that the industrial process still looks like a black box, which is not conducive to process understanding for operators. Visual process monitoring is the application of a visualization method to map the industrial process data to a two-dimensional map, followed by process monitoring. Visualization provides powerful insight and process understanding in the industrial context, and accelerates fault diagnosis [9]. Self-organizing map (SOM) is widely used for data visualization because it can retain the topological relations and density distribution of a data set when projecting the high-dimensional data to the

two-dimensional space. In addition, the common visualization methods include PCA, and some manifold learning methods, such as isometric feature mapping (ISOMAP) [10], Laplacian Eigenmaps (LE) [11], and t-distributed stochastic neighbor embedding (t-SNE) [12]. SOM has the following advantages in visualization compared with PCA and manifold learning methods. First, SOM is a nonlinear method based on the neural network. PCA is a linear dimensionality reduction method and cannot cope with the nonlinear data. Second, the output of SOM is 2-dimensional, which is specially designed for data visualization. However, the manifold learning methods and PCA have their essential dimensions for dimensionality reduction. They may cause loss of information and may not correctly reflect the structure of the high-dimensional data when they are used for data visualization. Finally, SOM does not have the out-of-sample problem. The out-of-sample problem is that the method does not have a mapping function when visualizing the training data, so that the method can not visualize the new sample. Some manifold learning methods, such as ISOMAP, LE, and t-SNE, have the out-of-sample problem, so they are not suitable for the classification task. Therefore, many studies have applied SOM to process monitoring [13–16].

The continuous development of modern science and technology has increased the equipment requirement of industrial systems. Massive variables are measured through sensors, which makes

* Corresponding author at: P.O. Box 293, MeiLong Road No. 130, Shanghai 200237, China. Fax: +86 21 64251036.

E-mail address: xfyan@ecust.edu.cn (X. Yan).

the high-value information difficult to observe directly. If these high-dimensional data are directly sent to the SOM for clustering, then the training complexity of the network will increase, and the training results will not reflect the true industrial operation states. This situation will result in poor visual process monitoring. The common practice to deal with the issue is to extract important features from the original data. Feng and Xu [17] used PCA to extract important features from the original data, and then used SOM to visualize these features for fault diagnosis. Chen and Yan [18,19] combined SOM with CCA and LDA to improve the performance of SOM for process monitoring. Similarly, Song *et al.* [20] combined SOM with CVA for process monitoring. Among these methods, the method combining SOM with LDA (LDA-SOM) can separate different classes relatively well and has relatively good performance for process monitoring. The objective of LDA is to ensure that the different classes are as far apart as possible while guaranteeing that the scatter of each class is as small as possible [21–24]. Therefore, the high-dimensional data from different classes can be separated as much as possible on the output of LDA-SOM. For visual process monitoring, the distances among classes should be as large as possible, and the samples from the same class should be as concentrated as possible. Large between-class distances and small within-class distances are conducive to improving the performance of visual process monitoring. However, one problem occurs in the application of LDA. The traditional LDA criterion overemphasizes the contribution of the classes far from the center in the between-class scatter measure. Thus, the neighboring classes usually overlap. Some works have been put forward to solve this problem with the corresponding algorithm. Loog *et al.* [25] proposed the approximate pairwise accuracy criterion (APAC). It expresses the between-class scatter matrix (BCSM) as the sum of the covariance matrix of the means of each class pair and assigns each class pair a weight function determined by the Mahalanobis distance between the class pair. [26] also introduced a weighting strategy based on class saliency information into LDA. Bian *et al.* [27] proposed a new criterion for feature extraction called max–min distance analysis (MMDA). It maximizes the minimum pairwise distance of all class pairs. Li *et al.* [28] developed the maximum margin criterion (MMC) for feature extraction by changing the criterion function of conventional LDA to $\text{tr}(S_b - S_w)$, where S_b is the BCSM and S_w is the within-class scatter matrix (WCSM). MMC ensures that the different classes are as far apart as possible while guaranteeing that the scatter of each class is as small as possible. Local Fisher discriminant analysis (LFDA) [29–31] is an extended form of LDA and is also a commonly used method for feature extraction.

In this paper, a novel criterion for discriminative feature extraction called balanced multiple weighted linear discriminant analysis (BMWLDA) is proposed to solve the problem of LDA over-optimizing the classes that are already well separated. Its basic ideas are as follows. BMWLDA maximizes the distances between neighboring classes while minimizing the scatter of neighboring classes. Thus, the classification result is in a balanced state, that is, different classes are completely separated. For the between-class scatter measure, BMWLDA removes the class pairs with large pairwise distances and then assigns a weight function that is negatively correlated with the pairwise distance for each remaining class pair. For the within-class scatter measure, BMWLDA removes classes far from all other classes and then assigns one weight function that is positively correlated with the scatter of the class to each remaining class. Therefore, BMWLDA can make a relatively uniform distribution of class centers and approximately maintain class separation.

Furthermore, we combine BMWLDA with SOM (BMWLDA-SOM) to visualize monitoring of industrial operation processes. BMWLDA is applied to extract the discriminative features from

the original industrial data and maximally separate industrial operation states in the space spanned by these discriminative feature vectors. When the discriminative feature vectors are used as the input to SOM, the training result of SOM can differentiate industrial operation states clearly. This feature greatly improves the performance of visual monitoring of these states. We also combine MMDA, APAC, LDA, LFDA, and MMC with SOM to obtain MMDA-SOM, APAC-SOM, LDA-SOM, LFDA-SOM, and MMC-SOM. BMWLDA-SOM is more capable of separating industrial operation states than LDA-SOM, APAC-SOM, MMC-SOM, MMDA-SOM, and LFDA-SOM through the simulation studies on Continuous stirred tank reactor (CSTR). Therefore, the feature extraction performance of WLDA is better than that of LDA, APAC, MMC, MMDA, and LFDA. CSTR is also applied to verify that BMWLDA-SOM can effectively perform visual process monitoring in real time.

The main contributions of the paper are summarized as follows:

- (1) BMWLDA is proposed for discriminative feature extraction. BMWLDA not only solves the problem of LDA over-optimizing the classes that are already well separated, but also improves the separation among classes.
- (2) The method that combines BMWLDA with SOM can effectively perform visual process monitoring in real time.

2. Self-Organizing Map

SOM is an unsupervised self-organizing algorithm that can automatically cluster input patterns [32,33]. It can retain the topological relations of data and project the samples with similar attributes to close locations on the two-dimensional plane. The architecture of SOM includes one input layer and one output layer. The neurons between two layers are full connected, and lateral suppression exists between the output neurons.

Assuming there are q k -dimensional input samples, then i th sample is $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ik}]^T$, where $i = 1, 2, \dots, q$. The weight parameter corresponding to the j th output neuron is $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jk}]^T$, and there are m output neurons. Kohonen algorithm is a representative learning method of the SOM network. The main steps are listed as follows [34].

- (1) Assign random numbers to the weight vectors of output neurons. Set iteration time $t = 0$, and the total number of iterations is T . Determine the value of learning rate $\eta(0)$, where $0 < \eta(0) < 1$.
- (2) Select one input pattern $\mathbf{x}_i = [x_{i1}(t), x_{i2}(t), \dots, x_{ik}(t)]^T$ to input into the input layer.
- (3) Calculate the Euclidean distance d_j between the j th output neuron and $\mathbf{x}_i$ as follows.

$$d_j = \|\mathbf{x}_i - \mathbf{w}_j\| = \sqrt{\sum_{n=1}^k (x_{in} - w_{jn}(t))^2} \quad (1)$$

- (4) Find the neuron b with the smallest Euclidean distance d_j and then call it the best-matching unit (BMU).

$$\|\mathbf{x}_i - \mathbf{w}_b\| = \min \{d_j\} \quad (2)$$

- (5) Update the weights of the BMU and its neighboring neurons according to the following formula.

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + \eta(t)h_{c,j}(t)(\mathbf{x}_i - \mathbf{w}_j(t)) \quad (3)$$

where $\eta(t)$ is a learning rate, and it is formulated as

$$\eta(t) = \eta(0) \exp\left(-\frac{t}{\lambda}\right) \quad (4)$$

where λ is a time constant. $h_{c,j}(t)$ is formulated as

$$h_{c,j}(t) = \exp\left(-\frac{d_{c,j}^2}{2\alpha^2(t)}\right) \quad (5)$$

where $d_{c,j}$ is the distance between the output neuron and the BMU. $\alpha(t)$ is the neighborhood width, and it is formulated as

$$\alpha(t) = \alpha(0) \exp\left(-\frac{t}{\lambda}\right) \quad (6)$$

(6) Select the next input pattern to input into the input layer and then return to step (3). Until all input patterns have been learned, go to step (7).

(7) Let $t = t + 1$ and then return to step (2). Until $t = T$, the training process is completed.

3. Balanced Multiple Weighted Linear Discriminant Analysis

3.1. Linear discriminant analysis

Supposing the input data contain c classes, and the number of input data is N , and each sample contains m variables. The class i has n_i samples. $\mathbf{x}_i^j$ represents the j th sample of class i .

The BCSM of LDA is

$$S_b = \sum_{i=1}^c n_i (\mathbf{z}_i - \mathbf{z}) (\mathbf{z}_i - \mathbf{z})^T \quad (7)$$

and the WSCM of LDA is

$$S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} (\mathbf{x}_i^j - \mathbf{z}_i) (\mathbf{x}_i^j - \mathbf{z}_i)^T \quad (8)$$

where $\mathbf{z}_i$ is the average corresponding to class i , and $\mathbf{z}$ represents the average of N samples.

The target of LDA is to solve the following formula.

$$\max_{\mathbf{W}} \frac{\mathbf{W}^T S_b \mathbf{W}}{\mathbf{W}^T S_w \mathbf{W}} \quad (9)$$

Eq. (9) can be solved using Eq. (10).

$$S_b \mathbf{W}'_q = \lambda_q S_w \mathbf{W}'_q \quad (10)$$

where λ_q is the eigenvalue corresponding to the eigenvector $\mathbf{W}'_q$. From the characteristic Eq. (10), the LDA projection matrix $\mathbf{W}_{\text{FDA}}$ can be obtained, which consists of the eigenvectors corresponding to the top $c-1$ eigenvalues obtained by eigenvalue decomposition of $(S_w)^{-1} S_b$. That is, $\mathbf{W}_{\text{FDA}} = [\mathbf{w}'_1, \mathbf{w}'_2, \dots, \mathbf{w}'_{c-1}]$.

One drawback of the LDA is that the traditional LDA criterion overemphasizes the contribution of the classes far from the center in the between-class scatter measure in multiple classifications, thereby resulting in overlapping of neighboring classes. Therefore, we propose a novel criterion BMWLDA to overcome this drawback.

3.2. Balanced multiple weighted linear discriminant analysis algorithm

3.2.1. New BCSM

The new BCSM uses the form proposed by Loog *et al.* [25], which is the sum of the covariance matrix of the means of each class pair as follows:

$$S'_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j (\mathbf{z}_i - \mathbf{z}_j) (\mathbf{z}_i - \mathbf{z}_j)^T \quad (11)$$

The class pairs with large pairwise distances are already fully separated. We should focus more on class pairs with small pairwise distances than the above-mentioned class pairs. We define a critical distance that is the average of pairwise distances of all class pairs. When the pairwise distance of the class pair is larger than the critical distance, this class pair is considered to have the large pairwise distance. Otherwise, this class pair is considered to have the small pairwise distance. c kinds of data and $c(c-1)/2$ class pairs are available. The critical distance is expressed as follows:

$$d_{\text{cri}} = \frac{2}{c(c-1)} \sum_{i=1}^{c-1} \sum_{j=i+1}^c d_{ij} = \frac{2}{c(c-1)} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \sqrt{(\mathbf{z}_i - \mathbf{z}_j)^T (\mathbf{z}_i - \mathbf{z}_j)} \quad (12)$$

where d_{ij} is the pairwise distance of classes i and j . BMWLDA removes the class pairs with large pairwise distances in the between-class scatter measure to completely separate the class pairs with small pairwise distances. Eq. (11) can be further rewritten as

$$S'_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^c l_{ij} n_i n_j (\mathbf{z}_i - \mathbf{z}_j) (\mathbf{z}_i - \mathbf{z}_j)^T, \quad l_{ij} = \begin{cases} 0, & \text{if } d_{ij} \geq d_{\text{cri}} \\ 1, & \text{otherwise} \end{cases} \quad (13)$$

Some pairwise distances of the remaining class pairs are relatively large, whereas others are relatively small. The class pairs with small pairwise distances may overlap after feature extraction. An intuitive idea is to assign a weight function $\bar{w}_1$ that is negatively correlated with the pairwise distance to each remaining class pair. The final BCSM is formulated as:

$$\bar{S}_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^c l_{ij} \bar{w}_1(d_{ij}) n_i n_j (\mathbf{z}_i - \mathbf{z}_j) (\mathbf{z}_i - \mathbf{z}_j)^T \quad (14)$$

where $\bar{w}_1(d_{ij})$ is the weight function and is expressed as

$$\bar{w}_1(d_{ij}) = \frac{1}{d_{ij}^2} = \frac{1}{(\mathbf{z}_i - \mathbf{z}_j)^T (\mathbf{z}_i - \mathbf{z}_j)} \quad (15)$$

3.2.2. New WSCM

The edge class is the class far from other classes. The scatter of this class can be ignored because it does not generally overlap with other classes. In this work, the class with the distance larger than d_{cri} from other classes is considered the edge class. BMWLDA removes edge classes in the within-class scatter measure. The WSCM is rewritten as

$$S'_w = \sum_{i=1}^c f_i \sum_{j=1}^{n_i} (\mathbf{x}_i^j - \mathbf{z}_i) (\mathbf{x}_i^j - \mathbf{z}_i)^T \quad (16)$$

where f_i is

$$f_i = \begin{cases} 0, & \text{if } \min \{d_{ij}\} > d_{\text{cri}}, j = 1, \dots, c \text{ and } j \neq i \\ 1, & \text{otherwise} \end{cases} \quad (17)$$

Some classes have relatively large scatter, whereas others have relatively small scatter. The scatter of class i is expressed as

$$S_i = \sqrt{\frac{1}{n_j} \sum_{j=1}^{n_j} (\mathbf{x}_i^j - \mathbf{z}_i)^T (\mathbf{x}_i^j - \mathbf{z}_i)} \quad (18)$$

The class with small scatter is still closely clustered after feature extraction. Thus, we should pay considerable attention to the class with large scatter. BMWLDA assigns one weight function that is

positively correlated with the scatter of the class to each remaining class in the within-class scatter measure. The final WCSM is expressed as

$$\bar{S}_w = \sum_{i=1}^c f_i \bar{w}_2(s_i) \sum_{j=1}^{n_i} (\mathbf{x}_i^j - \mathbf{z}_i) (\mathbf{x}_i^j - \mathbf{z}_i)^T \quad (19)$$

where $\bar{w}_2(s_i)$ is the weight function and is expressed as

$$\bar{w}_2(s_i) = s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{x}_i^j - \mathbf{z}_i)^T (\mathbf{x}_i^j - \mathbf{z}_i) \quad (20)$$

The BMWLDA criterion has the same form as the LDA criterion. The objective of BMWLDA is to solve the following formula.

$$\max_{\mathbf{v}} \frac{\mathbf{v}^T \bar{S}_b \mathbf{v}}{\mathbf{v}^T \bar{S}_w \mathbf{v}} \quad (21)$$

where $\mathbf{v}$ is the projection vector. Let $\mathbf{v}^T \bar{S}_w \mathbf{v} = 1$, the objective is transformed into solving the following problem.

$$\begin{cases} \max_{\mathbf{v}} \mathbf{v}^T \bar{S}_b \mathbf{v} \\ \text{s.t. } \mathbf{v}^T \bar{S}_w \mathbf{v} = 1 \end{cases} \quad (22)$$

According to the Lagrange multiplier method, there is

$$L(\mathbf{v}, \alpha) = \mathbf{v}^T \bar{S}_b \mathbf{v} + \alpha(1 - \mathbf{v}^T \bar{S}_w \mathbf{v}) \quad (23)$$

where α is the Lagrange multiplier.

Let $\partial L / \partial \mathbf{v} = 0$, there is

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{v}} &= 2(\bar{S}_b \mathbf{v} - \alpha \bar{S}_w \mathbf{v}) = 0 \\ \Rightarrow \bar{S}_b \mathbf{v} &= \alpha \bar{S}_w \mathbf{v} \end{aligned} \quad (24)$$

Therefore, the optimal projection vectors are equal to the generalized eigenvectors of the eigenvalue problem:

$$(\bar{S}_w)^{-1} \bar{S}_b \mathbf{v} = \alpha \mathbf{v} \quad (25)$$

The rank of $\bar{S}_b$ does not exceed $c-1$, so the number of eigenvectors of $(\bar{S}_w)^{-1} \bar{S}_b$ does not exceed $c-1$. Therefore, the eigenvectors corresponding to the top $c-1$ eigenvalues of $(\bar{S}_w)^{-1} \bar{S}_b$ are used to construct the projection matrix $\mathbf{V}_{\text{BMWLDA}}$. The feature vector t_i of the data point x_i is expressed as

$$t_i = \mathbf{V}_{\text{BMWLDA}}^T x_i \quad (26)$$

4. Self-Organizing Map Integrated with Balanced Multiple Weighted Linear Discriminant Analysis

BMWLDA-SOM includes offline modeling and online monitoring for visual process monitoring. In offline modeling, BMWLDA extracts the discriminative features from the training data and maximally separates industrial operation states in the space spanned by these discriminative feature vectors. Then, the discriminative features are used to train SOM. In online monitoring, the new features are extracted from the new data collected from the industrial process by using the projection matrix $\mathbf{V}_{\text{BMWLDA}}$ attained in offline modeling. The steps of visual process monitoring are listed as follows:

4.1. Offline modeling

- (1) Collect the historical data and then standardize them as $\mathbf{X}_{\text{train}}$.
- (2) Calculate the matrices $\bar{S}_b$ and $\bar{S}_w$ by using Eqs. (14) and (19). Subsequently, calculate the projection matrix $\mathbf{V}_{\text{BMWLDA}}$.

- (3) Project $\mathbf{X}_{\text{train}}$ onto $\mathbf{V}_{\text{BMWLDA}}$ to get the training features $\mathbf{T}_{\text{train}}$ by using Eq. (26).
- (4) Train SOM with $\mathbf{T}_{\text{train}}$ following the algorithm steps in Section 2. After training, different classes are divided into their respective areas, and the trained BMWLDA-SOM can be used for process monitoring.

4.2. Online monitoring

- (1) Obtain the data and then normalize them as the new data $\mathbf{X}_{\text{new}}$.
- (2) Project $\mathbf{X}_{\text{new}}$ onto $\mathbf{V}_{\text{BMWLDA}}$ to obtain the new features $\mathbf{T}_{\text{new}}$ by using Eq. (26).
- (3) Send $\mathbf{T}_{\text{new}}$ to the trained SOM. According to the area where the mapped point falls, we can monitor the operation states of the industrial process.

5. Experimental Result and Analysis

5.1. Continuous stirred tank reactor

Experimental data generated by simulation on a non-constant temperature process called CSTR [35–37]. The simple irreversible exothermic reaction $A \rightarrow B$ reacts in the reactor. Some of the heat must be removed by the coolant flowing through the jacket to control the temperature of the reactor and the composition of the reaction product. The structure of CSTR is shown in Fig. 1, and Table 1 shows the physical meaning of each variable in Fig. 1.

The following assumptions must be made when creating the CSTR model.

- (1) Keep the total volume of the material before and after the reaction in the reactor constant.
- (2) Keep the materials in the reactor evenly mixed.
- (3) Retain the parameter values throughout the reaction.

Under the above-mentioned assumptions, a mathematical model based on the law of conservation of energy is established as follows [35].

$$\frac{dC_A}{dt} = -k_0 e^{-E/(RT)} C_A + \frac{Q_F C_{AF} - Q_C C_A}{Ah} \quad (27)$$

$$\frac{dT}{dt} = \frac{k_0 e^{-E/(RT)} C_A (-\Delta H)}{\rho C_p} + \frac{Q_F T_F - Q_C T}{Ah} + \frac{UA_c (T_c - T)}{\rho C_p Ah} \quad (28)$$

$$\frac{dT_c}{dt} = \frac{Q_c (T_{CF} - T_c)}{V_c} + \frac{UA_c (T_c - T)}{\rho_c C_{pc} V_c} \quad (29)$$

$$\frac{dh}{dt} = \frac{Q_F - Q}{A} \quad (30)$$

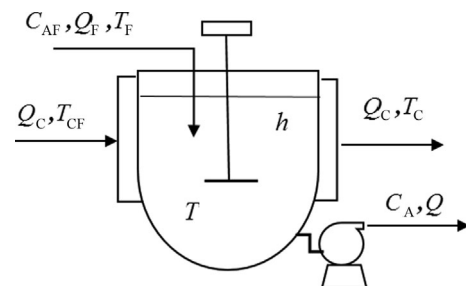


Fig. 1. Structure of CSTR.

Table 1

Variable description.

Variable	Description	Variable	Description
C_{AF}	Feed concentration	Q_F	Feed flow
T_F	Feed temperature	Q_C	Coolant flow
T_{CF}	Coolant inlet temperature	T_C	Coolant outlet temperature
C_A	Component A concentration in the reactor	Q	Component A Outlet flow
T	Temperature in the reactor	h	Liquid height

The meaning of parameters of the CSTR model is as follows: k_0 represents the reaction rate constant, E/R represents the activation energy coefficient, S represents the bottom area of the reactor; ΔH represents the reaction heat, ρ represents the reactant density, U_{Ac} represents the heat transfer coefficient, C_p represents the reactant specific heat, C_{pC} represents the specific heat of coolant, and V_C represents the volume of coolant. In the CSTR process, the 10 variables C_A , T , T_C , h , Q_F , Q , C_{AF} , Q_C , T_F , and T_{CF} are selected as monitoring variables. The steady operation conditions are listed in Table 2. Table 3 shows eight faults, including step and slow drift faults. The sampling interval is 0.01 min, and 600 samples are collected. The fault is introduced from the 301st sample.

5.2. Fault diagnosis based on SOM

First, we use training data to explore the ability of SOM to separate different operation states. Then, we use test data to test its fault diagnosis ability. 300 training and 300 test samples are selected for each state generated from the CSTR process.

5.2.1. Research on normal state and fault 1

The SOM network is trained using the training data of normal state and fault 1. The training result is shown in Fig. 2(a). On the output map, “N” indicates the label of the training sample of the normal state, “F1” indicates the label of the training sample of fault 1, and the number in the parentheses indicates the number of training samples mapped into the corresponding grid. In addition, different colors represent different kinds of data. The larger the color shape in the grid, the more samples are mapped into the grid. For example, “F1(5)” indicates that five training samples of fault 1 are mapped into the grid. The red represents the normal state, and the yellow represents fault 1. Fig. 2(a) shows that the two kinds of data are divided into two areas with clear boundaries. The SOM network has the capability of fault diagnosis when it is completely trained. Then, the performance of fault diagnosis of the trained SOM is tested. The test result is shown in Fig. 2(b). Two kinds of test data are mapped onto the correct area, and no sample is divided into the wrong class. The accuracy of fault diagnosis is 100%.

Table 2

Steady operation conditions of CSTR.

Variable	Standard value	Variable	Standard value
Q	100 L·min ⁻¹	A	0.1666 m ²
Q_C	15 L·min ⁻¹	k_0	$7.2 \times 10^{10} \text{ min}^{-1}$
T_F	320 K	ΔH	$-5 \times 10^4 \text{ J·mol}^{-1}$
T_{CF}	300 K	ρC_p	$239 \text{ J·L}^{-1}·\text{K}^{-1}$
T	402.35 K	$\rho_C C_{pC}$	$4175 \text{ J·L}^{-1}·\text{K}^{-1}$
C_{AF}	1.0 mol·L ⁻¹	U_{Ac}	$5 \text{ J·min}^{-1}·\text{K}^{-1}$
T_C	345.44 K	E/R	8750 K
C_A	0.037 mol·L ⁻¹	V_C	10 L
h	0.6 m	Q_F	100 L·min ⁻¹

Table 3

Eight kinds of fault description.

Fault no.	Fault description	Fault value
1	U_{Ac} slow drift	6 J·min ⁻¹ ·K ⁻¹
2	Q_F step change	4 L
3	T_{CF} slow drift	0.01 K·min ⁻¹
4	E/R slow drift	0.1 K·min ⁻¹
5	C_{AF} slow drift	$1.2 \times 10^{-4} \text{ mol·L}^{-1}·\text{min}^{-1}$
6	T measurement slow drift	0.4 K·min ⁻¹
7	T setpoint step change	2 K
8	T_F slow drift	0.015 K·min ⁻¹

5.2.2. Research on normal state and faults 1 and 2

Normal state and faults 1 and 2 are selected for further exploring the fault diagnosis capability of SOM. Fig. 3 shows that fault 1 is divided into its own area and can be completely separated from other states, whereas the normal state and fault 2 are mixed together. Therefore, we use BMWLDA to extract the discriminative features from the original data for fully separating different states, and then SOM is applied to visualize these features. This method is expected to separate different states on the output plane, and BMWLDA-SOM is researched next for fault diagnosis.

5.3. Fault diagnosis based on BMWLDA-SOM

5.3.1. Research on normal state and faults 1 and 2

Normal state and faults 1 and 2 are still selected to evaluate BMWLDA-SOM. BMWLDA is used to extract the discriminative features, and then these features are sent to train SOM. The training result is shown in Fig. 4(a). The three kinds of data are clustered into three different areas with clear boundaries. The test result is shown in Fig. 4(b). No sample is mapped into the wrong classes, and the accuracy of fault diagnosis is 100%. In BMWLDA-SOM, BMWLDA can extract the discriminative features that improve the separation between classes. Therefore, the fault diagnosis capability of BMWLDA-SOM is better than that of SOM.

5.3.2. Research on normal state and faults 1, 2, 3, 4, 5, 6, 7, and 8

Normal state and faults 1, 2, 3, 4, 5, 6, 7, and 8 are selected for further exploring the fault diagnosis capability of BMWLDA-SOM. The training result is shown in Fig. 5(a). The nine classes are divided into nine different areas. The test result is shown in Fig. 5(b). The normal state has three samples that are mapped into fault 1, and its accuracy is 99%. The accuracy of other classes is 100%. Although the number of fault classes has increased, BMWLDA-SOM still has a high accuracy of fault diagnosis. The experiment proves that the fault diagnosis performance of BMWLDA-SOM is effective.

5.4. Comparison of LDA, APAC, MMC, MMDA, LFDA, and BMWLDA for class separation

APAC, MMDA, MMC, and LFDA are improved feature extraction methods based on LDA. Since SOM can project the data into low-dimensional space and maintain the topological structure and density distribution of the input data, SOM is used to visualize the class separation performance based on different feature extraction methods. Similar to BMWLDA-SOM, we combine LDA, APAC, MMC, MMDA, and LFDA with SOM to obtain LDA-SOM, APAC-SOM, MMC-SOM, MMDA-SOM, and LFDA-SOM. Normal state and faults 1, 2, 3, 4, 5, 6, 7, and 8 are selected to evaluate the six methods. The results based on the six methods are shown in Fig. 6(a)–(f).

Here, an overlap rate is used to evaluate the performance of visual results based on different methods. The overlap rate of each class is equal to the sample amount in the class that overlaps with

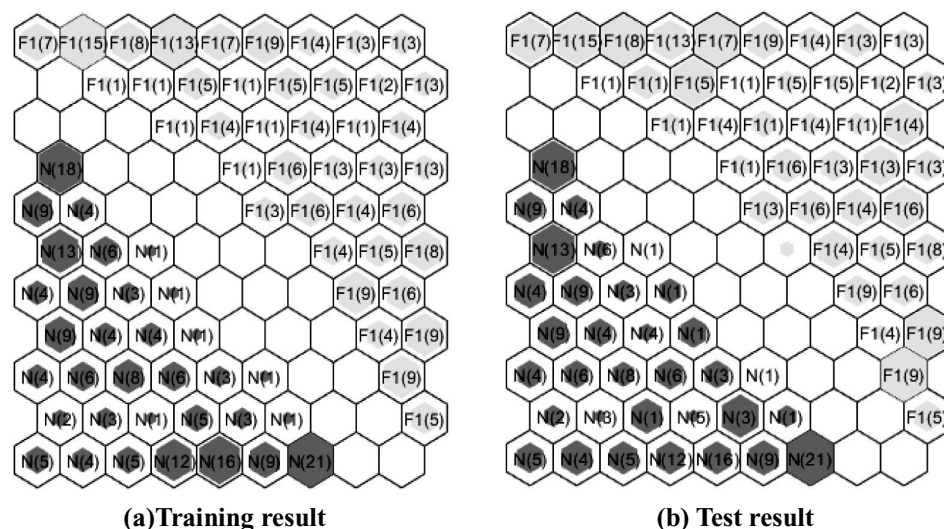


Fig. 2. Results for normal state and fault 1 based on SOM.

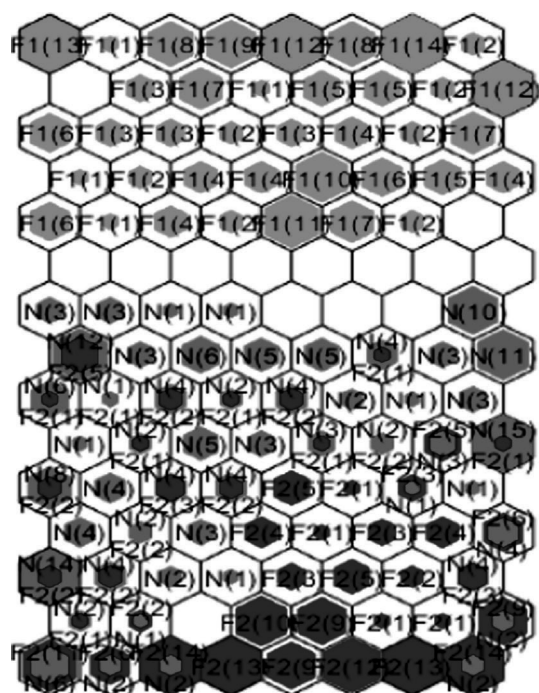


Fig. 3. Training result for normal state and faults 1 and 2 based on SOM.

samples of other classes divided by the total sample amount of the class. Furthermore, the average overlap rate is equal to the average of the overlap rates of all classes. The overlap rate for each method is listed in Table 4.

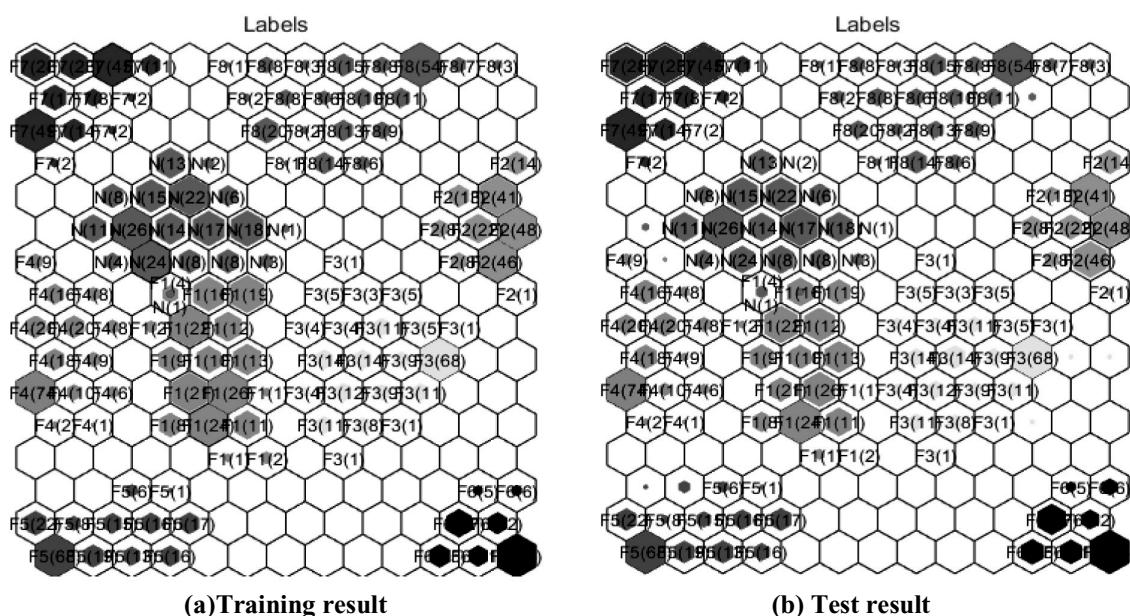
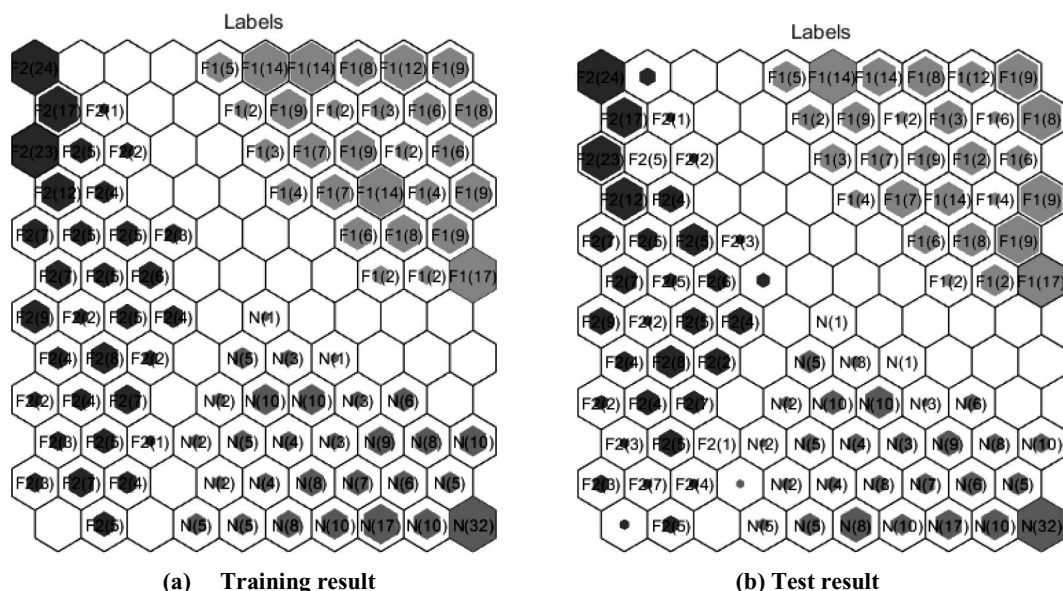
Faults 2 and 6 are quite different from normal state, and they are edge classes. The traditional LDA criterion overemphasizes the contribution of the classes far from the center in the between-class scatter measure, which easily leads to overlapping of neighboring classes. When LDA is used for feature extraction, normal state, faults 3, 4, and 8 are mixed together. APAC gives each class pair a weighting function in the between-class scatter measure. In APAC, if the Mahalanobis distance of the class pair is large, then the weight will be small. When APAC is used for feature extraction, fault 3 and normal state overlap significantly, but other states are nearly separated without overlapping. MMDA maxi-

mizes the minimum pairwise distance of all class pairs. When MMDA is used for feature extraction, normal state and faults 3 and 8 have a large overlap, whereas other states are nearly completely separated. MMC changes the objective function of traditional LDA to $\text{tr}(S_b - S_w)$. Its training result is much better than that of LDA, and only normal state and fault 3 have a large overlap, whereas other states are nearly completely separated. When LFDA is used for feature extraction, the training result is also much better than that of LDA, and only normal state and fault 1 have a large overlap, whereas other states are nearly completely separated.

BMW LDA maximizes the distances between neighboring classes while minimizing the scatter of neighboring classes. Thus, the classification result is in a balanced state, that is, different classes are fully separated. In the between-class scatter measure, BMW LDA removes the class pairs with large pairwise distances and then assigns a weight function that is negatively correlated with the pairwise distance for each remaining class pair. In the within-class scatter measure, BMW LDA removes classes far from all other classes and then assigns one weight function that is positively correlated with the scatter of the class to each remaining class. When BMW LDA is used for feature extraction, the nine classes are completely separated, and only a few samples overlap. The experiment shows that the class separation performance of BMW LDA is better than that of LDA, APAC, MMDA, MMC, and LFDA in multiple classifications. Therefore, BMW LDA-SOM is more suitable for fault diagnosis than LDA-SOM, APAC-SOM, MMDA-SOM, MMC-SOM, and LFDA-SOM.

5.5. Online monitoring based on BMW LDA-SOM

BMW LDA-SOM have the capacity to separate operation states of the industrial process with relatively clear boundaries according to the experiments above. This excellent visual result can greatly improve the accuracy and intuitiveness of process monitoring. When the industrial data are continuously fed into the trained BMW LDA-SOM, their mapped points are connected into a trajectory, and we can monitor the operation state of the industrial process according to the direction of the trajectory. We select the normal state and faults 2 and 3 for simulation. The entire process lasts 3 min. The data in the first 1 min are fault-free. Fault 2 is introduced after 1 min, and fault 3 is introduced after 2 min. The trajectory of monitoring is shown in Fig. 7. Fig. 7(a) shows that, in the first 1 min, the trajectory has been in the area of normal



average FDRs of PCA and KPCA are lower than those of other methods.

6. Conclusions

The traditional LDA criterion overemphasizes the contribution of the classes far from the center in the between-class scatter measure, thereby resulting in overlapping of neighboring classes. In this paper, BMWLDA is proposed to overcome the shortcomings of traditional LDA mentioned. Furthermore, BMWLDA-SOM is proposed to visualize monitoring of industrial operation process. CSTR is used to verify that the class separation performance of BMWLDA is better than that of LDA, APAC, MMC, MMDA, and LFDA. CSTR is also applied to illustrate that BMWLDA-SOM can effectively perform visual process monitoring in real time. However, there are

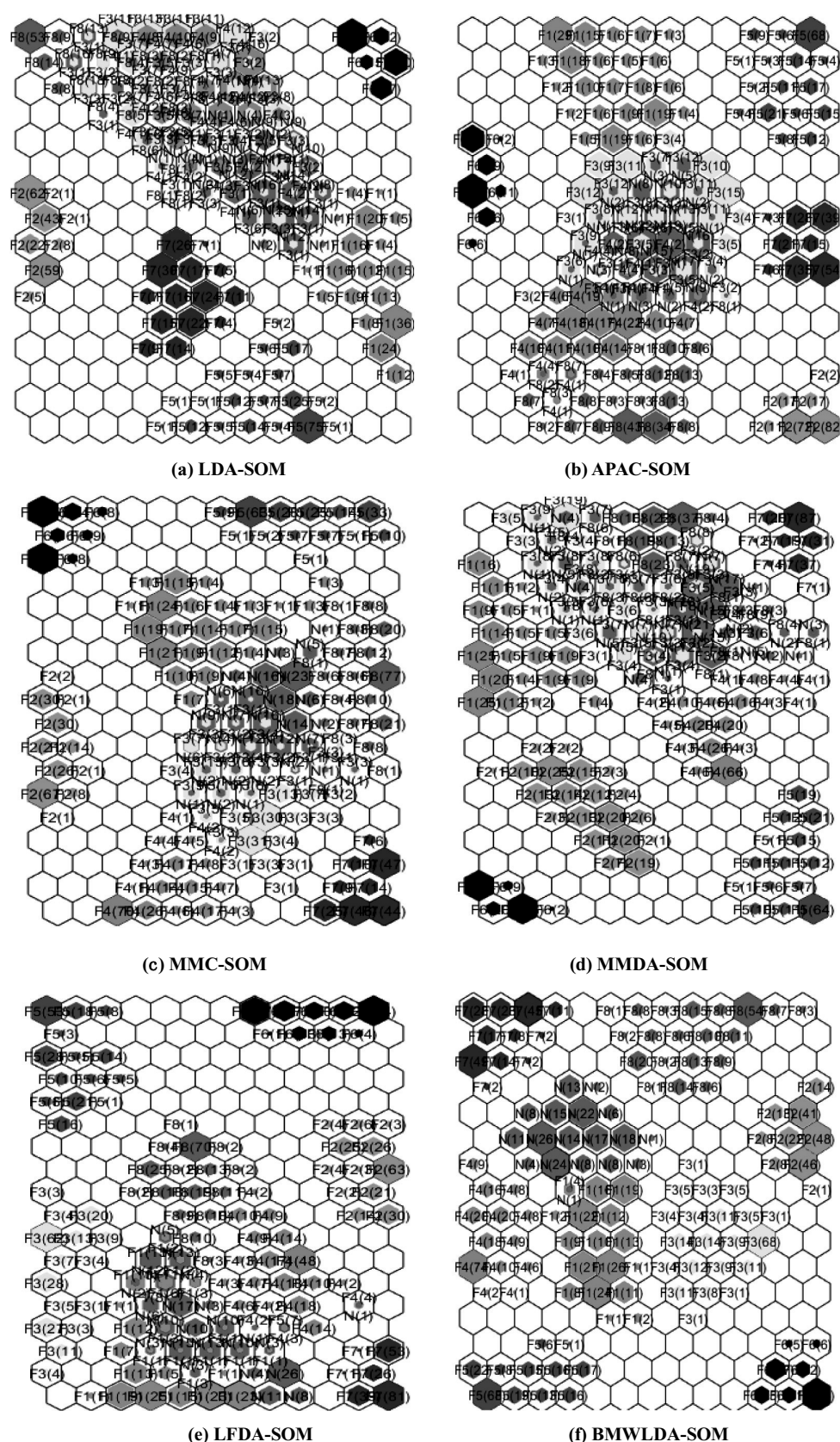


Fig. 6. Visual results of different methods.

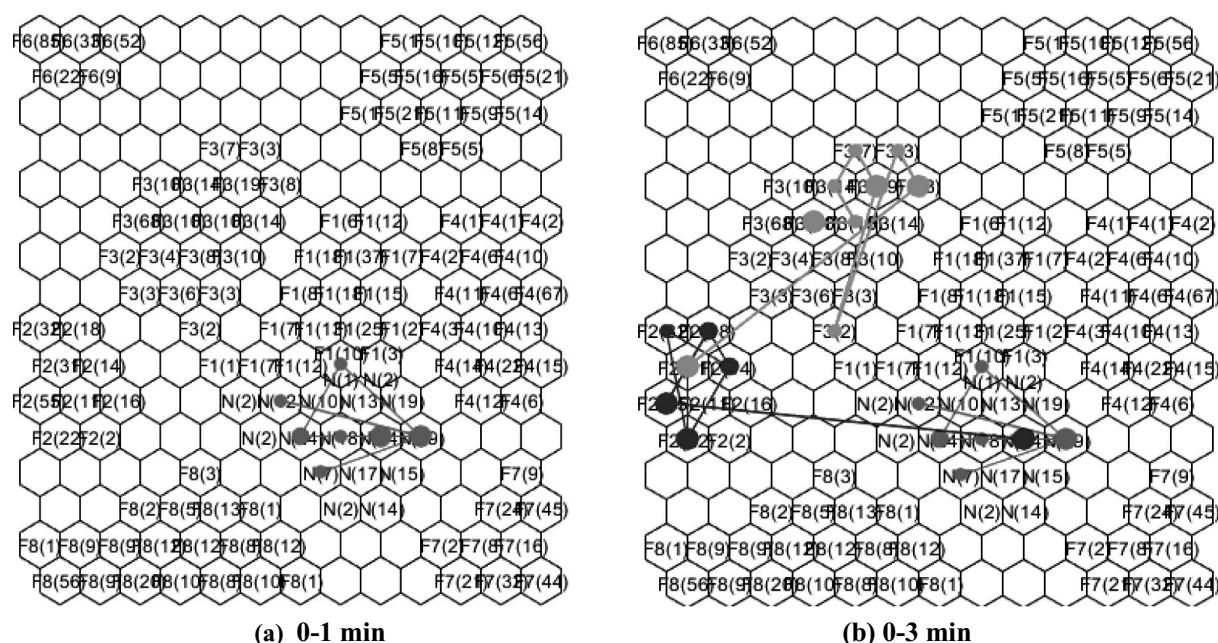
complex nonlinear relationships among variables in the industrial process. BMWLDA is essentially a linear method and can not extract the nonlinear information in the industrial process, so it

can not separate some complex faults well. The future work will extend BMWLDA to a non-linear version to improve the performance of process monitoring.

Table 4

Comparison of overlap rates of different methods with that of BMWLDA-SOM

Fault No.	LDA-SOM	APAC-SOM	MMC-SOM	MMDA-SOM	LFDA-SOM	BMWLDA-SOM
Normal	0.886	1	0.567	0.886	0.607	0.005
Fault 1	0	0	0	0	0.278	0.020
Fault 2	0	0	0	0	0	0
Fault 3	0.965	0.637	0.463	0.8756	0	0
Fault 4	1	0.244	0.020	0	0.045	0
Fault 5	0	0	0	0	0.035	0
Fault 6	0	0	0	0	0	0
Fault 7	0	0	0	0	0	0
Fault 8	0.587	0.0647	0.025	0.338	0	0
Average	0.382	0.216	0.119	0.233	0.107	0.003

**Fig. 7.** Trajectory of monitoring based on BMWLDA-SOM.**Table 5**

Comparison of FDRs of different methods with that of BMWLDA-SOM

Fault No.	LDA-SOM	APAC-SOM	MMC-SOM	MMDA-SOM	LFDA-SOM	BMWLDA-SOM	PCA	KPCA
Normal	0.4367	0.4433	0.4967	0.4633	0.5233	0.9533	0.3567	0.4267
Fault 1	0.9867	0.9900	0.7467	0.9900	0.7633	0.9567	0.6133	0.7467
Fault 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9267	0.9600
Fault 3	0.4400	0.4067	0.6733	0.4633	0.9733	0.9800	0.4733	0.5033
Fault 4	0.60333	0.8633	0.9233	0.9700	0.9300	0.9767	0.5867	0.6033
Fault 5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9933	0.9567	0.9633
Fault 6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Fault 7	1.0000	1.0000	1.0000	1.0000	0.9933	1.0000	0.9833	0.9867
Fault 8	0.6267	0.8933	0.8967	0.5633	0.9233	0.9933	0.6333	0.6433
Average	0.7874	0.8441	0.8596	0.8252	0.9007	0.9837	0.7256	0.7559

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful for the support of National Key Research and Development Program of China (2020YFA0908303), and National Natural Science Foundation of China (21878081).

References

- [1] X. Wang, U. Kruger, G.W. Irwin, Process monitoring approach using fast moving window PCA, *Ind. Eng. Chem. Res.* 44 (2005) 5691–5702.
- [2] G. Li, S.J. Qin, D.H. Zhou, Geometric properties of partial least squares for process monitoring, *Automatica* 46 (2010) 204–210.
- [3] J.M. Lee, C.K. Yoo, I.B. Lee, Statistical process monitoring with independent component analysis, *J. Process Contr.* 14 (2004) 467–485.
- [4] X.F. Yuan, J. Zhou, B. Huang, Y.L. Wang, C.H. Yang, W.H. Gui, Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy, *IEEE T. Ind. Inform.* 16 (2020) 3721–3730.
- [5] X.F. Yuan, C. Ou, Y.L. Wang, C.H. Yang, W.H. Gui, A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process, *IEEE Trans. Neural Netw. Learn. Syst.* 251 (2019) 113382.

- [6] Z.P. Zhang, J.S. Zhao, A deep belief network based fault diagnosis model for complex chemical processes, *Comput. Chem. Eng.* 107 (2017) 395–407.
- [7] H. Wu, J.S. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, *Comput. Chem. Eng.* 115 (2018) 185–197.
- [8] X.F. Yuan, L. Li, Y. Shardt, Y.L. Wang, C.H. Yang, Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development, *IEEE T. Ind. Electron.* 68 (5) (2021) 4404–4414.
- [9] M. Joswiak, Y. Peng, I. Castillo, L.H. Chiang, Dimensionality reduction for visualizing industrial chemical process data, *Control Eng. Pract.* 93 (2019) 104189.
- [10] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 290 (2000) 2319–2323.
- [11] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (2003) 1373–1396.
- [12] L.V.D. Maaten, Learning a parametric embedding by preserving local structure, in: International Conference on Artificial Intelligence and Statistics, 2009, 384–391.
- [13] A.A. Cuadrado, I. Diaz, A.B. Diez, F. Obeso, J.A. Gonzalez, Visual data mining and monitoring in steel processes, In: 37th IAS Annual Meeting, Pittsburgh, PA, USA (2002) 493–500.
- [14] H.Y. Yu, F. Khan, V. Garaniya, A. Ahmad, Self-organizing map based fault diagnosis technique for non-Gaussian processes, *Ind. Eng. Chem. Res.* 53 (2014) 8831–8843.
- [15] G. Robertson, M.C. Thomas, J.A. Romagnoli, Topological preservation techniques for nonlinear process monitoring, *Comput. Chem. Eng.* 76 (2015) 1–16.
- [16] H.L. Garcia, L.M. Gonzalez, Self-organizing map and clustering for wastewater treatment monitoring, *Eng. Appl. Artif. Intel.* 17 (2004) 215–225.
- [17] Z.G. Feng, T. Xu, Comparison of SOM and PCA-SOM in fault diagnosis of ground-testing bed, *Procedia Engineer* 15 (2011) 1271–1276.
- [18] X.Y. Chen, X.F. Yan, Using improved self-organizing map for fault diagnosis in chemical industry process, *Chem. Eng. Res. Des.* 90 (2012) 2262–2277.
- [19] X.Y. Chen, X.F. Yan, Fault diagnosis in chemical process based on self-organizing map integrated with fisher discriminant analysis, *Chinese J. Chem. Eng.* 21 (2013) 382–387.
- [20] Y. Song, Q.C. Jiang, X.F. Yan, M.J. Guo, A multi-SOM with canonical variate analysis for chemical process monitoring and fault diagnosis, *J. Chem. Eng. Jpn.* 47 (2014) 40–51.
- [21] K. Eberhardt, C. Beleites, S. Marthandan, C. Matthäus, S. Diekmann, J. Popp, Raman and infrared spectroscopy distinguish replicative senescent from proliferating primary human fibroblast cells by detecting spectral differences mainly due to biomolecular alterations, *Anal. Chem.* 89 (2017) 2937–2947.
- [22] N.P. Chien, L.K. Lautz, Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity, *Sci. Total Environ.* 618 (2017) 379–387.
- [23] H. Jeong, Y. Ohno, Symmetric lifting posture recognition of skilled experts with linear discriminant analysis by center-of-pressure velocity, *Intel. Serv. Robot.* 10 (2017) 323–332.
- [24] J. Portillo-Portillo, R. Leyva, V. Sanchez, G. Sanchez-Perez, H. Perez-Meana, J. Olivares-Mercado, K. Toscano-Medina, M. Nakano-Miyatake, A view-invariant gait recognition algorithm based on a joint-direct linear discriminant analysis, *Appl. Intelligence* 48 (2017) 1200–1217.
- [25] M. Loog, R.P.W. Duin, R. Haeb-Umbach, Multiclass linear dimension reduction by weighted pairwise Fisher criteria, *IEEE T. Pattern Anal.* 23 (2001) 762–766.
- [26] L. Xu, A. Iosifidis, M. Gabbouj, Weighted linear discriminant analysis based on class saliency information, In: 2018 25th IEEE International Conferences on Image Processing, Athens, Greece (2018) 2306–2310.
- [27] W. Bian, D. Tao, Max-Min distance analysis by using sequential SDP relaxation for dimension reduction, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2011) 1037–1050.
- [28] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, *IEEE Trans. Neural Networks* 17 (2006) 157–165.
- [29] S. Kumar, M.K. Bhuyan, B.C. Lovell, Y. Iwahori, Hierarchical uncorrelated multiview discriminant locality preserving projection for multiview facial expression recognition, *J. Vis. Commun. Image Represent.* 54 (2018) 171–181.
- [30] L. Zhuo, B. Cheng, J. Zhang, A comparative study of dimensionality reduction methods for large-scale image retrieval, *Neurocomputing* 141 (2014) 202–210.
- [31] H. Ali, M. Hariharan, S. Yaacob, A.H. Adom, Facial emotion recognition using empirical mode decomposition, *Expert Syst. Appl.* 42 (2015) 1261–1277.
- [32] A. Ortiz, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, Improving MR brain image segmentation using self-organizing maps and entropy-gradient clustering, *Inf. Sci.* 262 (2017) 117–136.
- [33] W.P. Tsai, S.P. Huang, S.T. Cheng, K.T. Shao, F.J. Chang, A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map, *Sci. Total Environ.* 579 (2017) 474–483.
- [34] G. Abaei, A. Selamat, H. Fujita, An empirical study based on semi-supervised hybrid self-organizing map for software fault prediction, *Knowl.-Based Syst.* 74 (2015) 28–39.
- [35] A. Singhal, D.E. Seborg, Pattern matching in multivariate time series databases using a moving-window approach, *Ind. Eng. Chem. Res.* 41 (2002) 3822–3838.
- [36] M.C. Johannesmeyer, A. Singhal, D.E. Seborg, Pattern matching in historical data, *AIChE J.* 48 (2002) 2022–2038.
- [37] X.G. Deng, X.M. Tian, Sparse kernel locality preserving projection and its application in nonlinear process fault detection, *Chinese J. Chem. Eng.* 21 (2013) 163–170.