ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

F-Norm-Based Soft LDA Algorithm for Fault Detection in Chemical Production Processes

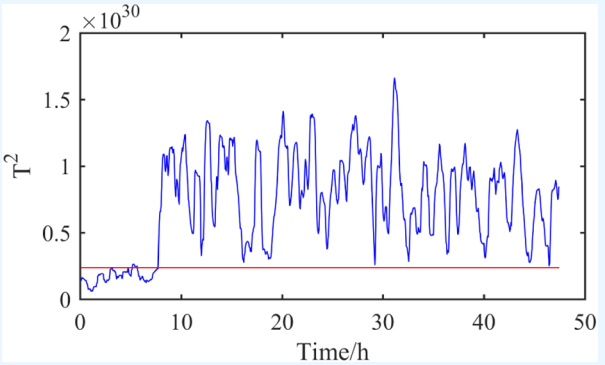
Hao Chen,* Haifei Zhang, Yuwei Yang, and Qiong Zhang

Cite This: *ACS Omega* 2024, 9, 34725−34734

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In chemical production processes, outliers are inevitable. Many existing feature extraction algorithms are overly sensitive to outliers and excessively focus on secondary features while ignoring the key features in the data. To address this problem, the Frobenius norm based soft linear discriminant analysis algorithm (FBSLA) is proposed in this paper. Specifically, FBSLA uses the Frobenius norm instead of its square as a metric to enhance the robustness of the algorithm. Furthermore, a nonreduced dimensionality projection matrix is introduced to make the training data features more obvious. Additionally, soft constraint is adopted instead of the traditional hard constraint to reduce the sensitivity caused by outliers. To validate the effectiveness of FBSLA, in this paper, experiments are conducted on the Tennessee Eastman Process and the Penicillin Fermentation Process data sets. According to experimental results, FBSLA significantly outperforms other state-of-the-art algorithms in terms of fault detection accuracy.



1. INTRODUCTION

The chemical production process is a complex system with many sensors and measuring devices monitoring various data. These data reflect the state of the production process, which is critical for ensuring proper equipment operation and increasing production efficiency.[1] However, during data monitoring of chemical processes, factors such as aging equipment, equipment fault, and environmental changes may lead to faulty data from measurement equipment. This can result in production process instability and an increase in operational risk.[2] Therefore, the study of fault detection in chemical production processes has become critical.

The rapid development of technology has driven the application of deep learning in fault diagnosis in chemical production processes.[3−5] For example, Yuan et al.[6] proposed a convolutional neural network model based on multiscale attention. The model generates feature maps containing multiscale features by employing convolution kernels of different sizes. Subsequently, by utilizing the channel attention mechanism, the model automatically emphasizes the importance of features at different scales, thereby enabling accurate prediction of industrial process quality. Similarly, Song and Jiang[7] designed a multiscale fault diagnosis method based on convolutional neural networks for high-dimensional and nonlinear chemical process data, which significantly improved fault detection accuracy. The performance of deep learning methods depends highly on the quality and quantity of data.[8] To ensure robustness and generalization, extensive and high-quality data sets are imperative for training. However, in practical chemical production processes, obtaining such data sets often poses significant challenges.

In contrast, machine learning has taken an important role in the field of chemical process fault diagnosis due to its interpretability and robustness. Machine learning methods usually show more adaptability and stability when facing outliers and noisy disturbances in the data. However, high-dimensional data often comes with redundant information and computational challenges. Therefore, effective dimensionality reduction techniques can distill key features from original data, revealing hidden patterns while reducing the dimensionality and complexity of data sets.[9] Many dimensionality reduction techniques, such as principal component analysis (PCA),[10] Kernel principal component analysis (KPCA),[11] and linear discriminant analysis (LDA),[12] have been widely used in diagnosing faults in the chemical production process. PCA can map high-dimensional data to a low-dimensional space and retain as much of the main features of the data as possible. However, PCA is sensitive to outliers, which limits its

Received: April 18, 2024
Revised: July 12, 2024
Accepted: July 25, 2024
Published: August 1, 2024

application in chemical production. To address this problem, improved PCA[13−15] enhances robustness by optimizing outlier description and selection mechanisms, increasing its reliability against outlier data. It is noteworthy that robust PCA aims to enhance data robustness. However, these methods may compromise data authenticity, leading to the loss of critical features. Additionally, robust PCA typically focuses less on category features, which poses a major limitation in chemical production. In chemical production processes, researchers focus on the differences between normal and abnormal processes, rather than differences within the data set.[16] To address this problem, LDA demonstrates its advantages. LDA is a supervised learning algorithm that effectively utilizes category features from both normal and abnormal process data. Therefore, LDA can capture and identify anomalous processes more accurately, allowing it to demonstrate greater effectiveness in chemical production fault diagnosis.

LDA enhances classification accuracy by maximizing intercategory differences, ensuring that data remains well separable after dimensionality reduction. However, traditional LDA has some limitations in data processing that can affect its performance under certain conditions. To overcome these limitations, researchers have proposed various improvement algorithms based on LDA, such as null-space LDA,[17] orthogonal LDA,[18] and ratio sum for linear discriminant analysis (RSLDA).[19] The null-space LDA maps the primitive feature space to a high-dimensional feature space, which can better capture the nonlinear structure of the data. However, the nonlinear mapping in the null-space LDA may amplify the effect of outliers. The orthogonal LDA and RSLDA enhance the classification performance by maximizing differences between categories. Nonetheless, these algorithms do not guarantee complete independence of the projected data from the original data. By optimizing the scatter matrix over the neighborhood, the limitations of traditional LDA under complex intraclass structures[20] can be addressed, thus describing the internal structure of the data more accurately. However, the above algorithms are still limited by hard constraints, making them sensitive to outliers and difficult to effectively capture the structural features of the data.

Using a sparse matrix is critical for efficiently representing the data and understanding its underlying structure. Sparse constraints improve the handling of high-dimensional data during dimensionality reduction. They also reduce the impact of outliers, enhancing the robustness of the algorithm.[21,22] For example, Hu et al.[23] addressed the problem of the squared L_2-norm exacerbating the influence of outliers when measuring model errors by introducing the L_1-norm to mitigate their impact. Similarly, Liang and Zhang[24] proposed a discriminant analysis algorithm based on the L_1-norm, enhancing the stability of LDA in the presence of outliers by studying the error bounds. Both above algorithms are based on the L_1-norm to increase the robustness of the algorithm. However, these algorithms are plagued by the nondeterministic polynomial-time hard problem. To overcome the limitations of the L_2-norm and L_1-norm, the L_{21}-norm is proposed.[25] The L_{21}-norm can better capture the overall distribution and shape of the data while remaining sparsity. Sha and Diao[26] proposed L_{21}-norm based KPCA to address the nonlinearity problem of data. Nie et al.[27] proposed a nongreedy strategy PCA based on the L_{21}-norm. Zhao et al.[28] proposed an improved LDA based on the L_{21}-norm, reducing the interference of outliers. Although the introduction of the L_{21}-norm has made progress in overcoming the limitations of the L_1 and L_2-norms, it remains necessary to address the dynamics of the data. Additionally, traditional LDA has not considered the dynamic features of the data over time when modeling, thus ignoring the changes in the data over the time series. Therefore, to better understand the dynamics of the data, dynamic processes[29] are introduced to capture the dynamic patterns over time.

Despite significant progress in the field of chemical production process fault detection, there are still many challenges. Deep learning excellence is highly dependent on the quantity and quality of data, which is particularly critical and difficult in the chemical industry. Traditional machine learning methods have shortcomings in dealing with outliers, and they often struggle to effectively focus on the dynamic and category features of data. Moreover, many existing methods are more sensitive to outliers due to hard constraints, making it difficult to fully and accurately capture the intrinsic structure of the data.

In summary, inspired by the L_{21}-norm, this paper uses the Frobenius norm (F-norm) instead of its square as a metric. Meanwhile, the interference of outliers is effectively reduced by using the soft constraint. Additionally, matrix samples are extracted using a sliding window and analyzed for dynamic processes. F-norm based soft LDA algorithm (FBSLA) is proposed in this paper, building upon these innovations and improvements. The main contributions of this paper are as follows:

1. To better retain the data change features, matrices are used as samples and extracted through a sliding window to reveal the dynamic features of the data. Meanwhile, to enhance the robustness of the algorithm, the F-norm is adopted instead of its square as the metric, accurately measuring the relationship between dimensions.

2. To make the features of the training data more obvious, a nonreduced dimensionality projection matrix is introduced to further sparse the data. This effectively reduces the dimensionality of the data while retaining key features, allowing the model to better learn and understand the underlying structure of the data.

3. The traditional hard constraint is replaced by a soft constraint to better accommodate changes in the data. A soft constraint can help to balance outliers and normal values in the data, making the model more robust when dealing with complex data.

The rest of the paper is organized as follows: Section 2 describes the detailed steps of the FBSLA. Section 3 describes the experimental procedure in detail. A discussion is presented in Section 4. Section 5 provides conclusions.

2. ALGORITHM

Traditional LDA typically does not account for changes in data over time, which limits their effectiveness in analyzing dynamic data sets. To address this problem, a dynamic LDA is adopted that allows the model to consider changes in data over time by introducing time series variables.[29] This paper mathematically models data sets containing time information to capture the temporal dynamics within the data. Specifically, the data from each batch in the database are first expanded into a two-dimensional data matrix using the time-lag window technique. The specific transformation equation can be expressed as follows

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

$$\mathbf{X}_u^i = \begin{bmatrix} (\mathbf{x}^i(u+1))^{\mathrm{T}} & (\mathbf{x}^i(u))^{\mathrm{T}} & \cdots & (\mathbf{x}^i(1))^{\mathrm{T}} \\ (\mathbf{x}^i(u+2))^{\mathrm{T}} & (\mathbf{x}^i(u+1))^{\mathrm{T}} & \cdots & (\mathbf{x}^i(2))^{\mathrm{T}} \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{x}^i(K))^{\mathrm{T}} & (\mathbf{x}^i(K-1))^{\mathrm{T}} & \cdots & (\mathbf{x}^i(K-u))^{\mathrm{T}} \end{bmatrix} \tag{1}$$

where K denotes the batch length, u denotes the time lag order, and i denotes the i-th batch.

The mathematical model of FBSLA is discussed next. Suppose the data has c categories, M_i denotes the number of samples in category c, $\mathbf{W} \in \mathrm{R}^{m\times k}$ denotes the projection matrix, $\overline{\mathbf{x}}_i$ denotes the mean of group i, $\overline{\mathbf{x}}$ denotes the average of all samples, and $\mathbf{x}_i^j$ denotes the j-th sample of group i. The objective function[17] of this optimal projection matrix can be expressed as follows

$$\operatorname{argmax} \frac{\sum_{i=1}^{c} M_i \| \mathbf{W}^{\mathrm{T}}(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}) \|_{\mathrm{F}}^2}{\sum_{i=1}^{c} \sum_{j=1}^{M_i} \| \mathbf{W}^{\mathrm{T}}(\mathbf{x}_i^j - \overline{\mathbf{x}}_i) \|_{\mathrm{F}}^2} \tag{2}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the F-norm of the matrix. In eq 2, the goal is to identify the optimal data projection direction by maximizing the distance between classes while minimizing the distance within each class. However, the presence of outliers may affect the estimates of intraclass and interclass variances, causing the projection directions to deviate from expectations.

In this paper, the MATLAB function is used to construct normal values and outliers to determine the optimal projection direction for LDA. Figure 1 illustrates the projection vectors of

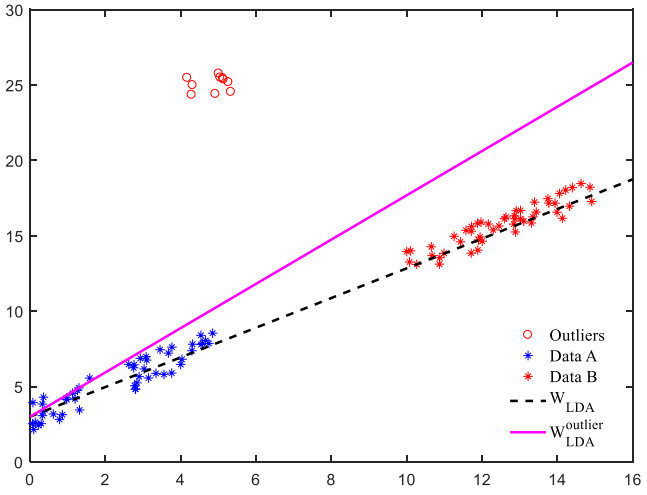


Figure 1. LDA projection vectors on a data set with and without outliers.

LDA on the data set with and without outliers. $\mathrm{W}_{\mathrm{LDA}}$ denotes the direction of the LDA projection without outliers, and $\mathrm{W}_{\mathrm{LDA}}^{\mathrm{outlier}}$ denotes the direction with outliers. From the figure, it can be observed that the presence of outliers causes a significant deviation between $\mathrm{W}_{\mathrm{LDA}}^{\mathrm{outlier}}$ and $\mathrm{W}_{\mathrm{LDA}}$. This deviation directly impacts the ability of the classifier to distinguish between different classes.

To minimize the negative impact of outliers on classification effects, one strategy involves replacing the squared term with a linear term in the objective function. The rationale is that the squared term amplifies the effect of distance, thus over-amplifying the influence of outliers far from the center point. In contrast, linear terms do not assign excessive weight to outliers far from the center, thereby helping to mitigate their impact on the overall objective function. Therefore, to enhance the robustness of the algorithm, this paper adopts the F-norm to measure both the intraclass and interclass scattering matrices within the objective function. Accordingly, the objective function is expressed as follows

$$\operatorname{argmax} \frac{\sum_{i=1}^{c} M_i \| \mathbf{W}^{\mathrm{T}}(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}) \|_{\mathrm{F}}}{\sum_{i=1}^{c} \sum_{j=1}^{M_i} \| \mathbf{W}^{\mathrm{T}}(\mathbf{x}_i^j - \overline{\mathbf{x}}_i) \|_{\mathrm{F}}} \tag{3}$$

When dealing with high-dimensional data, traditional dimensionality reduction strategies may encounter the problem known as dimensionality catastrophe. To address this problem, a novel projection matrix is introduced that does not reduce dimensionality in the conventional sense. Instead, it focuses on enhancing the separability of categories within the original feature space. This method clarifies category boundaries in the training data without directly reducing its dimensionality. Additionally, the sparsity of the data is further enhanced by using the L_{21}-norm. Building on this concept, the objective function is further defined as follows

$$\operatorname{argmax} \frac{\sum_{i=1}^{c} M_i \| \mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}) \|_{\mathrm{F}}}{\sum_{i=1}^{c} \sum_{j=1}^{M_i} \| \mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\mathbf{x}_i^j - \overline{\mathbf{x}}_i) \|_{\mathrm{F}} + \sum_{i=1}^{n} \| \mathbf{Q}^{\mathrm{T}}\mathbf{x}_i \|_{21}} \tag{4}$$

where $\mathbf{Q} \in R^{m\times m}$ denotes the nonreduced dimensionality projection matrix, and n denotes the number of samples.

To better explore potential associations in multidimensional data and preserve the original features, soft constraint is introduced.[30] Here a new matrix $\tilde{\mathbf{W}}_{(j)} \in R^{m\times(k-1)}$ is introduced, which includes the feature vectors other than $\mathbf{w}_j$. Suppose $\mathbf{w}_j$ can be expressed as follows

$$\mathbf{w}_j = d_{j1}\mathbf{w}_1 + d_{j2}\mathbf{w}_2 + \cdots + d_{j(k-1)}\mathbf{w}_{k-1} + d_{jk}\mathbf{w}_k = \tilde{\mathbf{W}}_{(j)}d_j \tag{5}$$

where d_j denotes the parameters of the feature vector. Next, minimize the correlation between $\mathbf{w}_j$ and $\tilde{\mathbf{W}}_{(j)}d_j$. Then define the following rule

$$\begin{aligned} R_{\mathrm{FF}} &= \sum_{j=1}^{k} r\langle \mathbf{w}_j,\ d_{j1}\mathbf{w}_1 + d_{j2}\mathbf{w}_2 + \cdots + d_{k-1}\mathbf{w}_{k-1} + d_{jk}\mathbf{w}_k \rangle \\ &= \sum_{j=1}^{k} r\langle \mathbf{w}_j,\ \tilde{\mathbf{W}}_{(j)}d_j \rangle = \frac{1}{t-1}\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}}) \end{aligned} \tag{6}$$

where $\tilde{\mathbf{W}} = [\tilde{\mathbf{W}}_1 d_1,\ \tilde{\mathbf{W}}_2 d_2, \cdots,\ \tilde{\mathbf{W}}_k d_k]$, $\mathrm{r}\langle \mathbf{a}, \mathbf{b} \rangle$ can be expressed as follows

$$r\langle \mathbf{a}, \mathbf{b} \rangle = \frac{1}{k-1} \sum_{i=1}^{k} \left(\frac{\mathbf{a}_i - \overline{\mathbf{a}}}{\sigma_{\mathrm{a}}} \right) \left(\frac{\mathbf{b}_i - \overline{\mathbf{b}}}{\sigma_{\mathrm{b}}} \right) = \frac{1}{k-1}\mathbf{a}^{\mathrm{T}}\mathbf{b} \tag{7}$$

where $\overline{\mathbf{a}}$ and $\overline{\mathbf{b}}$ are the mean values of $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively. σ_{a} and σ_{b} denote the standard deviation. The final simplified expression is shown as follows

$$R(\mathbf{W}) = \mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}}) \tag{8}$$

After adding the above soft constraint, the final objective function representation is obtained as follows

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

$$\begin{aligned}\operatorname{argmin} & \sum_{i=1}^{c}\sum_{j=1}^{M_i}\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)\|_{\mathrm{F}}+\lambda\sum_{i=1}^{n}\|\mathbf{Q}^{\mathrm{T}}\mathbf{x}_i\|_{21}\\ &-\eta\sum_{i=1}^{c}M_i\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})\|_{\mathrm{F}}+\mu_1\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}})\\ &+\mu_2\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\tilde{\mathbf{Q}})\end{aligned} \tag{9}$$

where λ, η, μ_1 and μ_2 denote the equilibrium parameters.

The above problem is unconstrained and can be solved by fixing $\mathbf{Q}$ and $\mathbf{W}$ respectively. The detailed solution steps are as follows

Step 1 (Update $\mathbf{W}$), fix $\mathbf{Q}$ first and then update $\mathbf{W}$. The update equation is as follows

$$\begin{aligned}\operatorname{argmin} & \sum_{i=1}^{c}\sum_{j=1}^{M_i}\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)\|_{\mathrm{F}}\\ &-\eta\sum_{i=1}^{c}M_i\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})\|_{\mathrm{F}}+\mu_1\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}})\end{aligned} \tag{10}$$

Then, eq 10 can be expressed as follows

$$\begin{aligned}\mathrm{tr}&\left(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\left(\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)D1(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)^{\mathrm{T}}\right)\mathbf{Q}\mathbf{W}\right)\\ &-\eta\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\sum_{i=1}^{c}M_i(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})D2(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})^{\mathrm{T}})\mathbf{Q}\mathbf{W})\\ &+\mu_1\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}})\end{aligned} \tag{11}$$

We let $A=\mathbf{Q}^{\mathrm{T}}\left(\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)D1(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)^{\mathrm{T}}\right)\mathbf{Q}$ and $B=\eta\mathbf{Q}^{\mathrm{T}}(\sum_{i=1}^{c}M_i(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})D2(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})^{\mathrm{T}})\mathbf{Q}$. Equation 11 can be expressed as follows

$$\mathrm{tr}(\mathbf{W}^{\mathrm{T}}A\mathbf{W})-\mathrm{tr}(\mathbf{W}^{\mathrm{T}}B\mathbf{W})+\mu_1\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}}) \tag{12}$$

Then, the following expression is obtained

$$\mathrm{tr}(\mathbf{W}^{\mathrm{T}}(A-B)\mathbf{W})+\mu_1\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{W}}) \tag{13}$$

The derivative of eq 13 with respect to $\mathbf{W}$ is given by

$$2(A-B)\mathbf{W}+\mu_1\tilde{\mathbf{W}}=0 \tag{14}$$

The final $\mathbf{W}$ can be expressed as follows

$$\mathbf{W}=-\frac{\mu_1}{2}(A-B)^{-1}\tilde{\mathbf{W}} \tag{15}$$

Step 2 (Update $\mathbf{Q}$), fix $\mathbf{W}$ first and then update $\mathbf{Q}$. The update equation is as follows

$$\begin{aligned}\operatorname{argmin} & \sum_{i=1}^{c}\sum_{j=1}^{M_i}\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)\|_{\mathrm{F}}+\lambda\sum_{i=1}^{n}\|\mathbf{Q}^{\mathrm{T}}\mathbf{x}_i\|_{21}\\ &-\eta\sum_{i=1}^{c}M_i\|\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})\|_{\mathrm{F}}+\mu_2\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\tilde{\mathbf{Q}})\end{aligned} \tag{16}$$

Then, eq 16 can be expressed as follows

$$\begin{aligned}&\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\left(\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)D1(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)^{\mathrm{T}}\right)\mathbf{Q}\mathbf{W})\\ &-\eta\mathrm{tr}\left(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\left(\sum_{i=1}^{c}M_i(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})D2(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})^{\mathrm{T}}\mathbf{Q}\mathbf{W}\right)\right)\\ &+\lambda\mathrm{tr}\left(\mathbf{Q}^{\mathrm{T}}\sum_{i=1}^{n}\mathbf{x}_iD_i\mathbf{x}_i^{\mathrm{T}}\mathbf{Q}\right)+\mu_2\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\tilde{\mathbf{Q}})\end{aligned} \tag{17}$$

We let $C=\lambda\mathbf{Q}^{\mathrm{T}}\sum_{i=1}^{n}\mathbf{x}_iD_i\mathbf{x}_i^{\mathrm{T}}\mathbf{Q}$, $D=\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)D1(\mathbf{x}_i^j-\overline{\mathbf{x}}_i)^{\mathrm{T}}$ and $E=\eta\sum_{i=1}^{c}M_i(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})D2(\overline{\mathbf{x}}_i-\overline{\mathbf{x}})^{\mathrm{T}}$. Equation 17 can be expressed as follows

$$\begin{aligned}&\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}D\mathbf{Q}\mathbf{W})-\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}E\mathbf{Q}\mathbf{W})+\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}C\mathbf{Q})\\ &+\mu_2\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\tilde{\mathbf{Q}})\end{aligned} \tag{18}$$

The derivative of eq 18 with respect to $\mathbf{Q}$ is given by

$$\Delta=2D\mathbf{Q}\mathbf{W}\mathbf{W}^{\mathrm{T}}-2E\mathbf{Q}\mathbf{W}\mathbf{W}^{\mathrm{T}}+2C\mathbf{Q}+\mu_2\tilde{\mathbf{Q}} \tag{19}$$

The solution to eq 19 obtained through gradient descent is given as follows:

$$\mathbf{Q}_{t+1}=\mathbf{Q}_t+\delta\Delta \tag{20}$$

where δ denotes the learning rate. Finally, by updating $\mathbf{Q}$ on a regular basis, the result is obtained.

Algorithm 1 is a summation of the previous formulas. It aims to compute two key matrices: the projection matrix $\mathbf{W}$ and the nonreduced dimensionality projection matrix $\mathbf{Q}$

Algorithm 1: F-norm Based Soft LDA Algorithm (FBSLA)

Input: Training data: $\mathbf{X}=[\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_{n-1},\mathbf{X}_n]$, parameter: λ, η, μ_1, μ_2.

Output: Projection matrix: $\mathbf{W}\in R^{m\times k}$, nonreduced dimensionality projection matrix: $\mathbf{Q}\in R^{m\times m}$.

1. Initialization parameters: λ, η, μ_1, μ_2;
2. While not converge do
3. Update $\mathbf{W}$ by Eq. (15);
4. Update $\tilde{\mathbf{W}}$ by Eq. (5) to Eq. (8);
5. Update $\mathbf{Q}$ by Eq. (20);
6. Update $\tilde{\mathbf{Q}}$ by Eq. (5) to Eq. (8).
7. End
8. Return $\mathbf{W}$, $\mathbf{Q}$, $\tilde{\mathbf{W}}$, $\tilde{\mathbf{Q}}$.

The T^2 statistic measures the difference between multiple variables. The difference between normal and fault states may be less obvious in the context of fault detection. Using T^2 can help detect such subtle differences. If the linear transformation does not change the distribution of the random variable, its covariance matrix[31] can be expressed as follows

$$\sum=(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{X})(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}} \tag{21}$$

Therefore, the final expression used for fault detection is as follows

$$T=n\times(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}\Sigma^{-1}(\mathbf{W}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{X}) \tag{22}$$

The kernel density estimation algorithm[32] is used to calculate the control limits of T^2.

3. EXPERIMENTS AND RESULTS

3.1. Datasets. The data in the Tennessee Eastman Process (TEP)[33] are obtained from 52 sensors, which are mainly used for fault diagnosis in chemical production processes. The data set includes training and validation sets, comprising samples of both

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

Table 1. Penicillin Data Generation Detail

state	fermentation period (h)	sampling time (h)	sample size	position to add disturbance	disturbance mode	disturbance value
normal	100	0.2	500			
fault 1	300	0.2	1500	200	step	−0.6
fault 2	300	0.2	1500	200	ramp	3
fault 3	300	0.2	1500	200	ramp	1

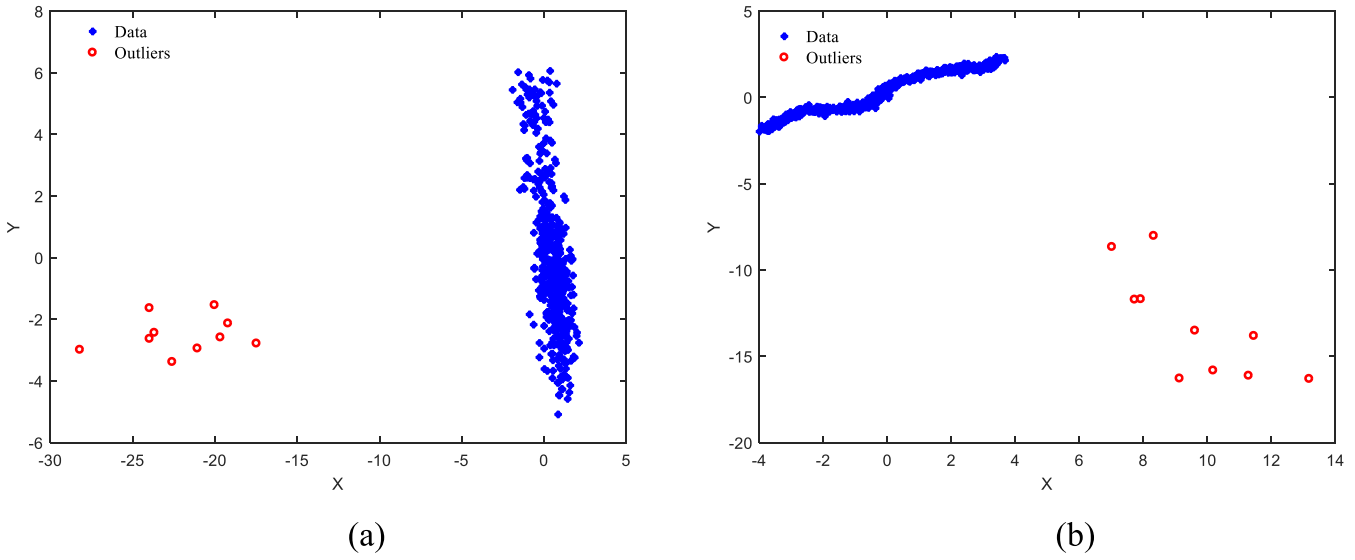


Figure 2. Scatter plot with added outliers. (a) TEP. (b) PFP.

normal and faulty instances. D00.dat to d21.dat are the training set samples, d00-te.dat to d21-te.dat are the test set samples, d00.dat and d00-te.dat are the samples under normal operation, and the others are the samples with faults. Thus, this data set contains 21 faults. Each training set sample with faults includes 480 observations and each test set sample includes 960 observations.

The penicillin fermentation process (PFP) data set is typically nonlinear in nature. The pensim simulation platform is utilized to generate data,[33] enabling fault detection within the penicillin production process. The data set covers normal operating conditions as well as fault 1, fault 2, and fault 3. This paper introduces two types of faults: step disturbance and ramp disturbance, which are applied to the aeration rate, agitator power, and substrate feed rate. Table 1 demonstrates the penicillin data generation in detail. The end point of the fermentation is set to 300 h and the sampling time is set to 0.2 h. All faults are added from 40 h until the end of the fermentation. Each training set sample with faults includes 500 observations and each test set sample includes 1000 observations.

To validate the effectiveness of FBSLA, 10 outliers are added to the TEP and PFP data sets, replacing the normal training data. This allows evaluation of the ability of the model to handle outliers. Figure 2a illustrates the scatter plot of the TEP data set with the 10 added outliers. Figure 2b illustrates the scatter plot of the PFP data set after adding 10 outliers. From the figures, it can be observed that the outliers are more prominent in the data distribution, which may cause interference to traditional LDA and PCA.

3.2. Parameter Setting. This paper selected seven values (0.001, 0.01, 0.1, 1, 10, 100, and 1000) for each of the four parameters (λ, η, μ_1, and μ_2) and performed parameter optimization experiments on fault 17 in TEP data set. Figure 3 shows the results of the parameter selection. The fault detection

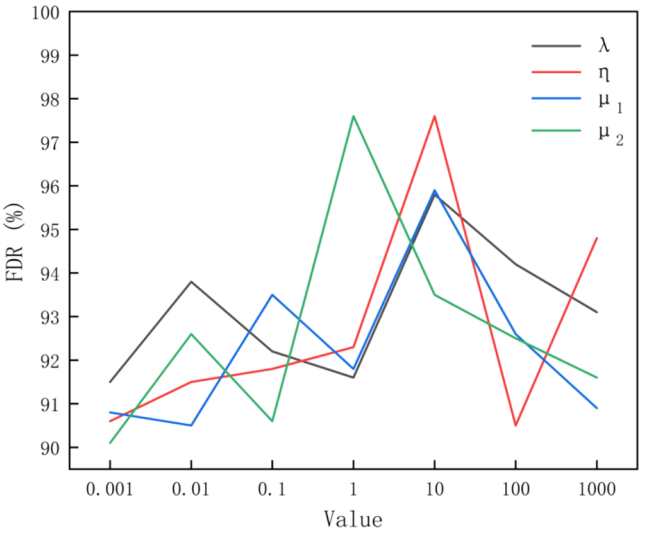


Figure 3. Impact of parameter values on FDP in fault 17.

rate (FDR) reaches its highest value when the values of λ, η, and μ_1 are 10, while it is highest when the value of μ_2 is 1. To further refine the parameters, this paper conducts more detailed experiments near these preliminary optimal values, and finally determines that the optimal parameters are $\lambda = 30$, $\eta = 10$, $\mu_1 = 30$, and $\mu_2 = 1$, respectively. Through this stepwise approximation method, the misclassification rate is reduced, providing a reliable basis for parameter selection.

3.3. TEP Experiment. The fault detection rate (FDR)[33] is a metric used to validate the performance of FBSLA. The expression to calculate the FDR is as follows

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

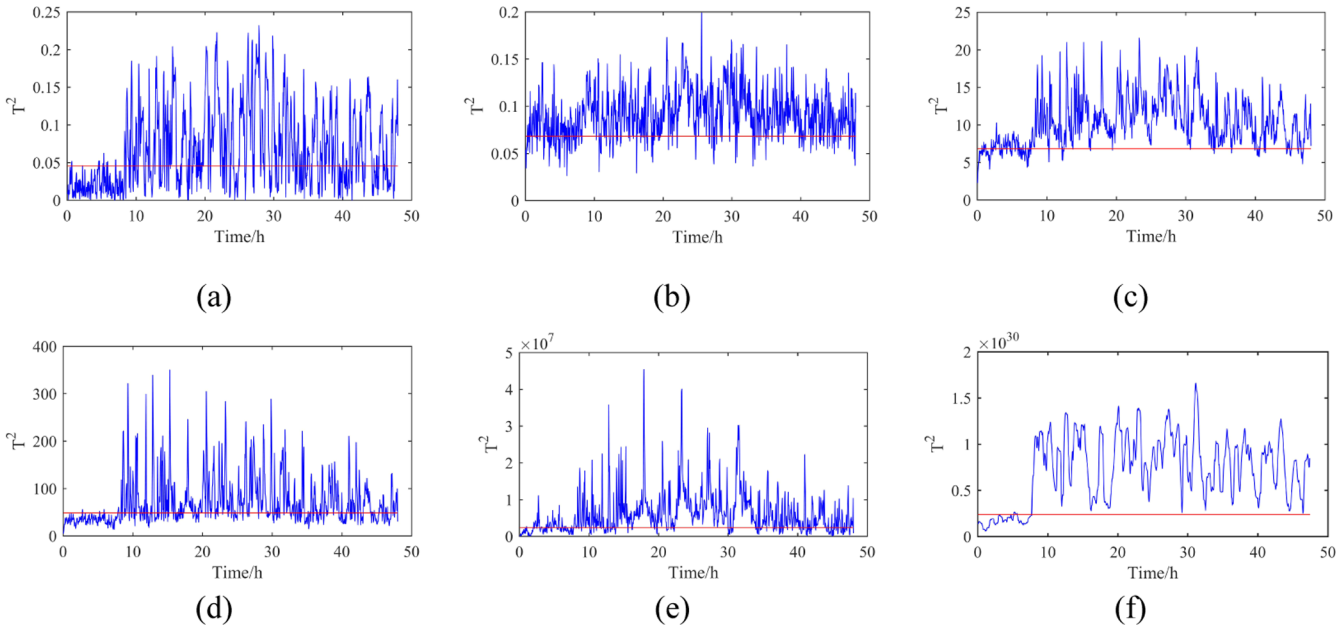


Figure 4. Detection results for fault 11. (a) LDA. (b) RLDA. (c) RSLDA. (d) PCA. (e) KPCA. (f) FBSLA.

$$\mathrm{FDR} = \frac{N_n + N_f}{N_s} \tag{23}$$

where N_n represents the number of successful detections in normal data, N_f represents the number of successful detections in fault data, and N_s represents the total number of test data. The higher the FDR, the better the algorithm is at detecting faults. To validate the effectiveness of the algorithm, FBSLA is experimentally compared with LDA,[12] RLDA,[28] RSLDA,[19] PCA,[10] and KPCA.[26]

Figure 4 shows the results of fault 11 detection. Figure 4a shows that the T^2 of the LDA begins to fluctuate around the control limit after 8 h. This fluctuation indicates that the LDA is having difficulty detecting faults consistently and effectively. Figure 4b displays the detection results of RLDA. Although RLDA improves its robustness by minimizing the L_{21}-norm, it still risks misjudging normal data as faults. In contrast, Figure 4c shows that RSLDA has an FDR of 82.4%. According to this figure, the majority of data is above the control limit after 8 h, indicating that RSLDA has a significant advantage in recognizing faults. RSLDA enhances the discriminative power of each feature, allowing for more effective utilization of interclass differences when handling high-dimensional data. In some cases, the algorithms in Figure 4d,e can detect faults, but their performance fluctuates, resulting in many missed faults. KPCA handles complex data by introducing kernel functions that map the data to a higher dimensional space. This makes KPCA outperform PCA in fault identification accuracy. However, selecting the kernel function for KPCA is more challenging, which directly impacts its performance in fault detection. Figure 4f shows that after 8 h, the statistics are higher than the control limit, indicating that FBSLA is the most effective algorithm. This is because FBSLA uses a nonreduced dimensionality projection matrix, which makes the features of the training data more prominent. This approach effectively captures the subtle differences in the data, thereby better preserving the original features and improving the accuracy of detection.

Table 2 illustrates the results of monitoring 21 faults in the TEP data set, with bold numbers indicating the best results under each set of faults. According to the results, FBSLA performs the best in most cases. By averaging the FDRs of the 21 fault cases, the average detection rate presented in the last column of Table 2 is obtained. Overall, in 21 sets of faults, the FDR outperforms the other algorithms. The average FDR for FBSLA is 83.0%, which is 12.9% higher than LDA, 10.8% higher than RLDA, 4.7% higher than RSLDA, 11.5% higher than PCA, and 3.3% higher than KPCA. However, for the detection of faults 9, 15, and 16, the FDR of FBSLA is less than 50%. This indicates that the algorithm is not as effective in detecting these specific faults.

To evaluate the sensitivity of different algorithms to outliers, this paper incorporates outliers as training data in the TEP data

Table 2. FDR (%) for the TEP Dataset

fault	LDA	RLDA	RSLDA	PCA	KPCA	FBSLA
1	98.6	81.6	90.8	98.4	97.2	**99.4**
2	96.8	81.6	92.2	98.2	96.0	**99.6**
3	29.0	50.2	61.8	26.2	**65.8**	50.9
4	**99.0**	81.0	82.0	79.8	75.2	98.8
5	68.4	81.4	**87.2**	44.4	69.0	84.5
6	99.4	85.8	92.8	99.1	97.9	**99.6**
7	98.6	84.4	90.2	**99.6**	73.6	94.1
8	82.4	77.6	81.2	97.0	90.4	**99.2**
9	31.6	55.8	53.6	24.2	**67.2**	37.7
10	63.8	77.2	79.0	62.9	77.6	**82.2**
11	75.2	70.6	82.4	71.2	77.4	**98.0**
12	77.2	84.2	86.4	**98.8**	92.6	98.4
13	88.6	84.0	92.2	96.2	96.2	**96.4**
14	37.6	66.0	87.4	**99.4**	88.8	99.2
15	34.0	53.4	50.4	29.7	**60.8**	32.7
16	55.0	**73.6**	67.8	46.7	72.6	42.7
17	93.8	79.4	81.0	88.2	85.8	**97.6**
18	82.8	82.0	82.8	91.6	92.3	**93.4**
19	30.4	50.6	67.6	36.9	53.5	**92.4**
20	73.4	74.4	82.4	61.5	82.4	**83.0**
21	56.2	40.4	52.6	52.5	61.2	**62.2**
mean	70.1	72.2	78.3	71.5	79.7	**83.0**

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

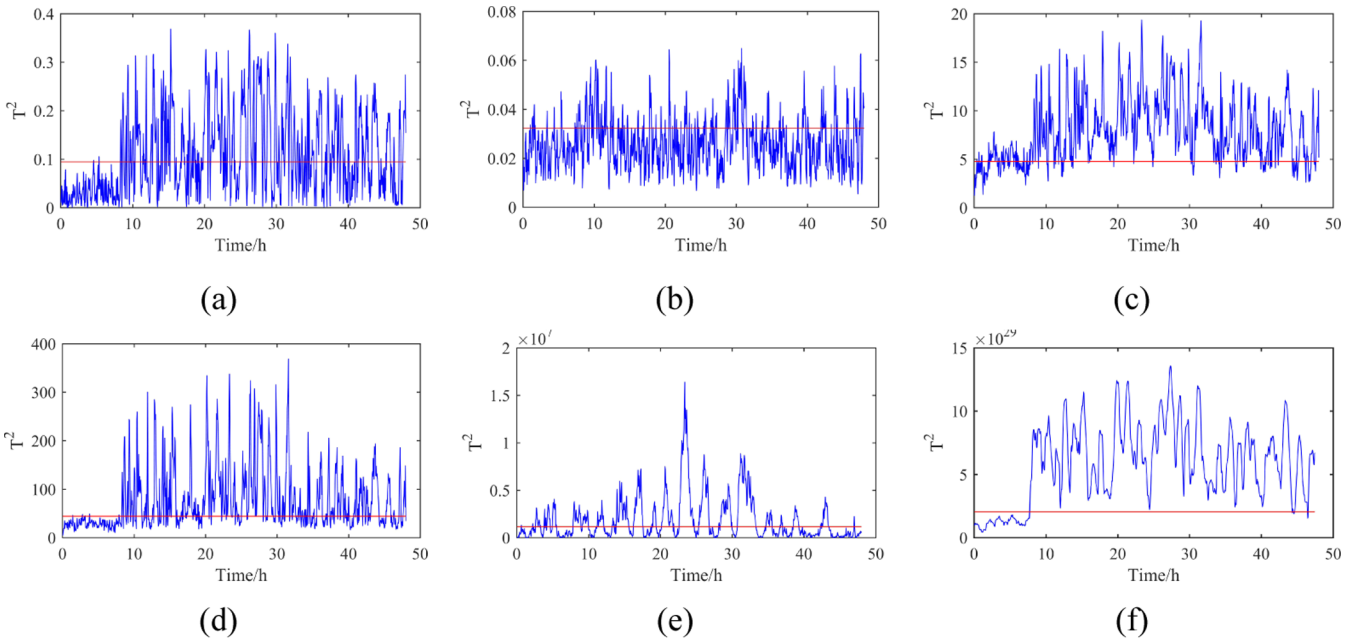


Figure 5. Detection results for fault 11 with outliers. (a) LDA. (b) RLDA. (c) RSLDA. (d) PCA. (e) KPCA. (f) FBSLA.

set. Figure 5a,f illustrates the detection results after adding outliers. When outliers are added, the fault detection results of LDA, RLDA, RSLDA, PCA, and KPCA remain poor. Although PCA and RSLDA are less affected by the outliers, the detection results are still poor. In contrast, FBSLA is least affected by outliers in the monitoring of fault 11. Before 8 h, the detected data always remains below the control limits, and after 8 h, it is almost always above them. This suggests that FBSLA still exhibits good fault detection with outliers.

Table 3 presents the monitoring results for the 21 faults with added outliers. It indicates that the FDR of FBSLA exceeds that of the other algorithms in 8 of the 21 sets of faults,

Table 3. FDR (%) for the TEP Dataset with Outliers

fault	LDA	RLDA	RSLDA	PCA	KPCA	FBSLA
1	98.9	94.5	95.5	99.3	83.0	**99.8**
2	96.9	95.1	93.1	98.3	57.6	**100.0**
3	29.6	32.0	**57.8**	21.8	50.4	51.0
4	98.9	99.0	79.8	**99.6**	49.1	99.2
5	41.7	**92.8**	89.6	40.6	61.0	90.0
6	99.4	96.3	94.1	**99.8**	97.3	98.6
7	98.3	99.0	89.6	**99.9**	64.6	99.0
8	81.9	44.3	82.9	**98.1**	86.4	91.7
9	29.6	40.0	**51.9**	20.2	50.2	36.7
10	47.2	37.8	**76.1**	57.2	65.8	69.8
11	29.6	40.4	81.2	71.0	59.2	**98.2**
12	79.4	72.0	87.8	**98.5**	87.5	87.5
13	89.6	69.2	92.6	95.8	91.4	**96.5**
14	29.6	36.5	90.0	99.4	55.4	**99.6**
15	29.6	36.3	45.7	27.3	**56.0**	33.5
16	29.6	45.1	**64.9**	43.3	59.5	48.2
17	82.8	48.6	80.9	80.5	74.0	**98.3**
18	83.5	80.4	83.1	90.9	89.6	**92.7**
19	29.6	37.1	**69.2**	30.9	41.4	34.7
20	68.7	83.3	85.3	56.7	71.8	**90.2**
21	49.8	38.0	51.8	53.5	**54.1**	52.8
mean	63.1	62.7	78.2	70.6	66.9	**79.4**

demonstrating its continued strong performance. The average FDR of FBSLA is 79.4%, compared to 63.1% for LDA, 62.7% for RLDA, 78.2% for RSLDA, 70.6% for PCA, and 66.9% for KPCA. These results further demonstrate the superiority of FBSLA in fault detection. It is noteworthy that RSLDA exhibits the least change with the addition of outliers compared to Table 2, indicating that it is the least affected by outliers. However, although RSLDA is less affected by outliers, its average detection accuracy is lower compared to FBSLA.

3.4. PFP Experiment. Figure 6 displays the results of detecting fault 1 in the PFP data set. As shown in Figure 6a,f, RLDA performs poorly with this data set, indicating that the algorithm may not be suitable. PCA detects the fault at 38 h, but the data from the subsequent assays fluctuate around the control limits. KPCA and RSLDA show similar detection results. In contrast, LDA does not detect the fault until 60 h. FBSLA detects the fault at 46 h, and the data after 46 h are all above the control limit, indicating that the fault is fully detected.

Table 4 demonstrates the performance of different algorithms in terms of FDR in the PFP data set, where FBSLA achieving the highest FDR for both faults. Additionally, FBSLA achieves an average FDR of 88.6%. Compared to LDA, FBSLA improves FDR by 12.1%; compared to RLDA, by 12.4%; compared to RSLDA, by 14%; compared to PCA, by 9.2%; and compared to KPCA, by 3.9%. These results further confirm the superiority of FBSLA in fault detection.

Figure 7a,f illustrates the detection results for fault 1 in the PFP data set. It is noteworthy that the detection effectiveness of all algorithms decreases to varying degrees, indicating that the introduction of outliers negatively affects the detection effect. Even though FBSLA is slightly affected in the detection of fault 1, it can still detect the majority of faults.

Table 5 shows the FDR of various algorithms after introducing outliers. FBSLA has the highest detection accuracy for faults 1 and 2. Furthermore, The FBSLA has the highest average FDR of the three faults. When comparing Table 4 with Table 5, it is observed that the addition of outliers results in a decrease in accuracy for all algorithms, with FBSLA experiencing

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

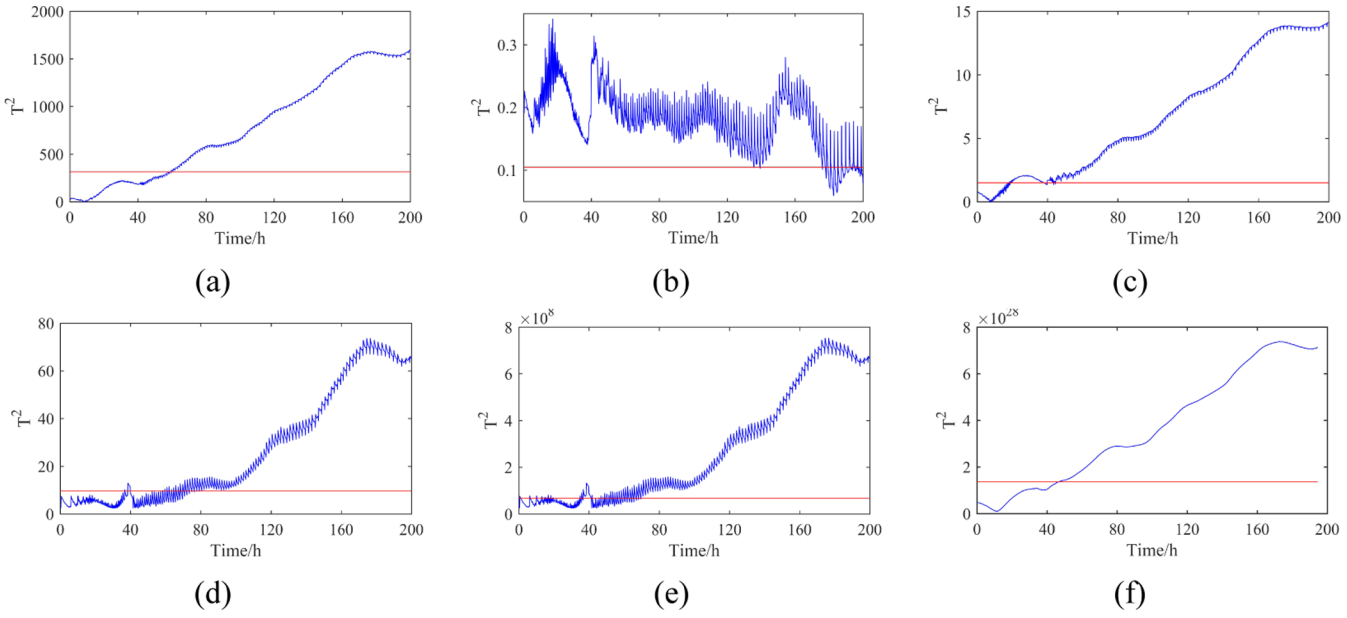


Figure 6. Detection results for fault 1. (a) LDA. (b) RLDA. (c) RSLDA. (d) PCA. (e) KPCA. (f) FBSLA.

Table 4. FDR (%) for the PFP Dataset

fault	LDA	RLDA	RSLDA	PCA	KPCA	FBSLA
1	77.0	78.0	79.8	75.2	87.5	**93.2**
2	77.0	78.0	73.8	**86.5**	81.4	85.2
3	75.6	72.6	70.2	76.5	85.3	**87.5**
mean	76.5	76.2	74.6	79.4	84.7	**88.6**

Table 5. FDR (%) for the PFP Dataset with Outliers

fault	LDA	RLDA	RSLDA	PCA	KPCA	FBSLA
1	42.1	77.3	73.7	53.1	69.5	**82.1**
2	79.6	44.5	67.6	82.8	83.3	**84.2**
3	75.5	74.7	55.5	**77.0**	70.7	76.6
mean	65.7	65.5	65.6	71.0	74.5	**81.0**

a 7.6% decrease in FDR. Compared to the other algorithms, FBSLA shows the smallest decrease in accuracy, confirming its superiority in dealing with outliers.

Table 6 illustrates the fault alarm rates for the TEP and PFP data sets under different conditions. This paper focuses on fault 11 (TEP) and fault 1 (PFP). For fault 11 of the TEP data set, the fault alarm rate is 100% without outliers. When outliers are added, the fault alarm rate decreases slightly to 99.3%, a decrease

Table 6. Comparison of Fault Alarm Rates (%) under Different Groups

group	fault alarm rate
TEP without outlier (fault 11)	100.0
TEP with outlier (fault 11)	99.3
PFP without outlier (fault 1)	87.8
PFP with outlier (fault 1)	85.6

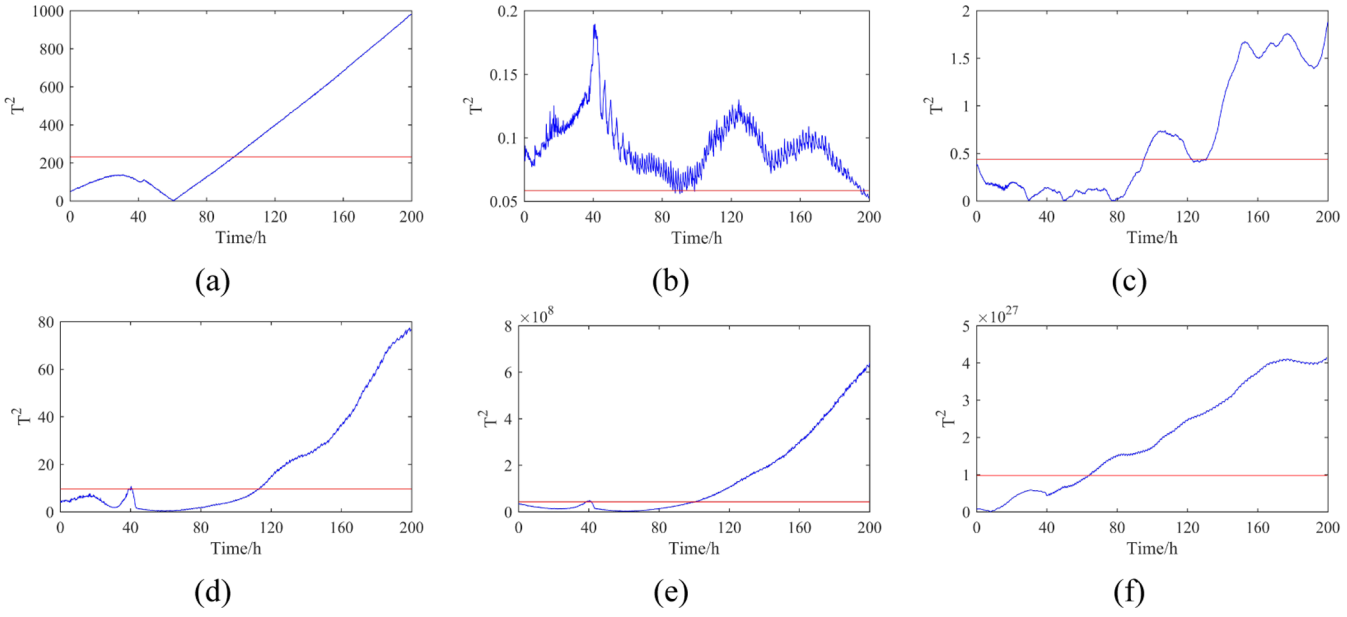


Figure 7. Detection results for fault 1 with outliers. (a) LDA. (b) RLDA. (c) RSLDA. (d) PCA. (e) KPCA. (f) FBSLA.

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

of 0.7%. For fault 1 of the PFP data set, the alarm rate is 87.8% without outliers and drops slightly to 85.6% with outliers added, resulting in a decrease of only 2.2%. Overall, the decrease in the fault alarm rate after outliers is minor, demonstrating that FBSLA maintains a high fault detection rate in the presence of outliers.

To deeply show the robustness of FBSLA, this paper keeps using Figure 1 as an example. Since matrices cannot be visualized in two dimensions, vectors are utilized in this paper instead of matrices. Figure 8 illustrates a comparison of the projection

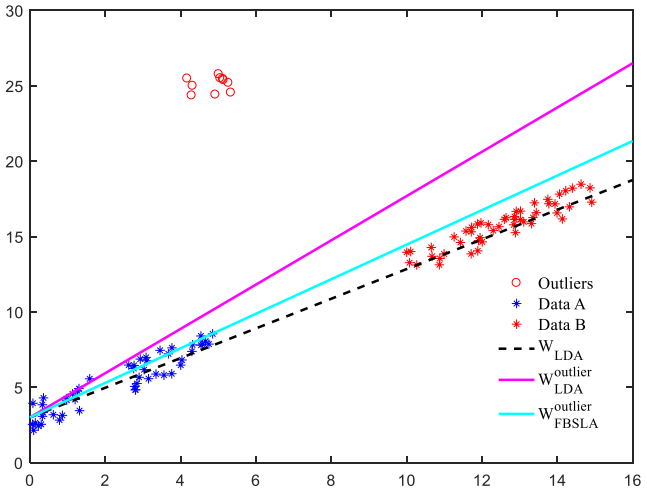


Figure 8. Comparison of projection vectors on a data set with and without outliers.

vectors of LDA and FBSLA, where W_{LDA} denotes the projection vector of LDA without outliers, $W_{LDA}^{outlier}$ denotes the projection vector of LDA with outliers, and $W_{FBSLA}^{outlier}$ denotes the projection vector of FBSLA with outliers. By comparing the distances of $W_{LDA}^{outlier}$ to W_{LDA}, and $W_{FBSLA}^{outlier}$ to W_{LDA}, it is clearly observed that when the data set contains outliers, $W_{FBSLA}^{outlier}$ and W_{LDA} remain highly close to each other, fully demonstrating the strong robustness of FBSLA in dealing with outliers.

4. DISCUSSION

LDA, based on Gaussian distribution assumptions, frequently encounters challenges related to robustness when applied to complex data sets. RLDA and RSLDA incorporate mechanisms to mitigate the impact of outliers on traditional LDA. However, these algorithms employ hard constraints, resulting in outliers being heavily weighted in the optimization process. In contrast, the application of soft constraints is an effective solution to reduce the negative impact of outliers. Soft constraints, unlike hard constraints, allow for a certain degree of suppression of outliers during the optimization process, reducing their impact on the results. This flexible strategy strengthens the algorithm in the face of outliers and reduces performance fluctuations. On the other hand, PCA is a widely used dimensionality reduction algorithm. It simplifies data and reduces computational complexity. However, its projection matrix may lose important features, which leads to certain key features being ignored in the dimensionality-reduced data. This can subsequently affect the accuracy of the analysis. KPCA is an enhanced version of PCA, which improves the performance to some extent, but still faces problems with robustness. When outliers exist in the data, KPCA may be affected, thereby limiting its applicability in certain scenarios.

In this paper, extensive experiments are conducted using the RLDA algorithm on the TEP and PFP data sets. However, the experimental results are not satisfactory for accurately detecting faults. Although the potential of RLDA to enhance algorithmic robustness has been recognized, its application to these specific data sets did not meet our expectations. This discrepancy suggests that RLDA is not suitable for fault detection in the TEP and PFP data sets.

FBSLA employs multiple innovations to address the problem. Specifically, it replaces the squared term with a linear term in the objective function. This enables FBSLA to effectively handle variations across different dimensions. Meanwhile, FBSLA employs a nonreduced dimensionality projection matrix. This strategy helps preserve the key features of the data. Therefore, it allows for a more accurate representation of the intrinsic structure of the data. Moreover, FBSLA innovatively employs soft constraints instead of hard constraints, significantly reducing the sensitivity of the data analysis process to outliers.

This paper reveals a remarkable finding by comparing the LDA and FBSLA projection vector results. It shows that the projection vectors of FBSLA with outliers are closer to the projection vectors of LDA without outliers. This not only demonstrates that FBSLA has better processing capability than LDA in dealing with outliers, but also emphasizes the high robustness of FBSLA. FBSLA demonstrates the ability to be less susceptible to interference when dealing with complex data structures, which is critical for handling outliers in practical problems. In chemical production processes, outliers often have a significant impact on the stability of the production process. FBSLA effectively addresses these outliers while preserving the intrinsic structure of the data set. In summary, FBSLA has demonstrated its effectiveness in the TEP and PFP data sets, maintaining stable performance in the presence of outliers.

5. CONCLUSIONS

FBSLA is proposed in this paper for the problem of fault detection in chemical production processes. FBSLA significantly enhances the robustness of the algorithm by introducing F-norm as a metric. Second, the nonreduced dimensionality projection matrix is added to make the training data features more obvious, thus retaining the key features in the data. Furthermore, traditional hard constraints are replaced by soft constraints. This substitution effectively reduces the interference outliers cause on algorithm performance. It also further enhances the robustness of FBSLA. Experimental validation on the TEP and PFP data sets shows that FBSLA outperforms other state-of-the-art algorithms. These strategies together enhance the accuracy of FBSLA in chemical production process fault detection.

However, it is critical to recognize that the algorithm has its own limitations, particularly when dealing with nonlinear problems that may be difficult. In the future, algorithms will continue to be optimized to better adapt to the complex and changing environments in chemical production. More accurate solutions for fault detection in chemical production processes can be provided by introducing kernel functions to solve nonlinear problems.

AUTHOR INFORMATION

Corresponding Author

Hao Chen − *School of Information Engineering, Nantong Institute of Technology, Nantong 226002, China;*

https://doi.org/10.1021/acsomega.4c03747
ACS Omega 2024, 9, 34725−34734

orcid.org/0009-0005-6986-4805; Email: chenhao@ntit.edu.cn

Authors

Haifei Zhang − *School of Information Engineering, Nantong Institute of Technology, Nantong 226002, China*

Yuwei Yang − *School of Information Engineering, Nantong Institute of Technology, Nantong 226002, China*

Qiong Zhang − *School of Information Engineering, Nantong Institute of Technology, Nantong 226002, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c03747

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was funded by Natural Science Foundation of University in Jiangsu Province (Grant 23KJD520011), Research topic on educational informatization in Jiangsu Higher Education Institutions (Grant 2023JSETKT126), Ministry of Education Industry-Education Cooperation Collaborative Education Project (Grant 220901046092840), and Nantong Natural Science Foundation (General Project) (Grant JC2023075). The authors are thankful to all the personnel who provided technical support. We also acknowledge all the reviewers for their useful comments and suggestions.

■ REFERENCES

(1) Qin, S. J. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* **2012**, *36* (2), 220−234.

(2) Md Nor, N.; Che, H. C.; Hussain, M. A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems. *Rev. Chem. Eng.* **2020**, *36* (4), 513−553.

(3) Yuan, X.; Xu, N.; Ye, L.; Wang, K.; Shen, F.; Wang, Y.; Yang, C.; Gui, W. Attention-based interval aided networks for data modeling of heterogeneous sampling sequences with missing values in process industry. *IEEE Trans. Ind. Inf.* **2024**, *20* (4), 5253−5262.

(4) Deng, L.; Zhang, Y.; Dai, Y.; Ji, X.; Zhou, L.; Dang, Y. Integrating feature optimization using a dynamic convolutional neural network for chemical process supervised fault classification. *Process Saf. Environ. Prot.* **2021**, *155*, 473−485.

(5) Yuan, X.; Wang, Y.; Wang, C.; Ye, L.; Wang, K.; Wang, Y.; Yang, C.; Gui, W.; Shen, F. Variable correlation analysis-based convolutional neural network for far topological feature extraction and industrial predictive modeling *IEEE Trans. Instrum. Meas.* 2024; Vol. *73* DOI: 10.1109/TIM.2024.3373085.

(6) Yuan, X.; Huang, L.; Ye, L.; Wang, Y.; Wang, K.; Yang, C.; Gui, W.; Shen, F. Quality prediction modeling for industrial processes using multiscale attention-based convolutional neural network. *IEEE Trans. Cybern.* **2024**, *54* (5), 2696−2707.

(7) Song, Q.; Jiang, P. A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes. *Process Saf. Environ. Prot.* **2022**, *159*, 575−584.

(8) Chen, X. W.; Lin, X. Big data deep learning: challenges and perspectives. *IEEE Access* **2014**, *2*, 514−525.

(9) Reddy, G. T.; Reddy, M. P. K.; Lakshmanna, K.; Kaluri, R.; Rajput, D. S.; Srivastava, G.; Baker, T. Analysis of dimensionality reduction techniques on big data. *IEEE Access* **2020**, *8*, 54776−54788.

(10) Zhu, J.; Ge, Z.; Song, Z. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Trans. Indus. Inform.* **2017**, *13* (4), 1877−1885.

(11) Lahdhiri, H.; Taouali, O. Reduced rank KPCA based on GLRT chart for sensor fault detection in nonlinear chemical process. *Measurement* **2021**, *169*, No. 108342.

(12) Wang, Y.; Wu, D.; Yuan, X. LDA-based deep transfer learning for fault diagnosis in industrial chemical processes. *Comput. Chem. Eng.* **2020**, *140*, No. 106964.

(13) Ahsan, M.; Mashuri, M.; Kuswanto, H.; Prastyo, D. D.; Khusna, H. Outlier detection using PCA mix based T^2 control chart for continuous and categorical data. *Commun. Stat.-Simul. Comput.* **2021**, *50* (5), 1496−1523.

(14) Wang, Y.; Yu, H.; Li, X. Efficient iterative dynamic kernel principal component analysis monitoring method for the batch process with super-large-scale data sets. *ACS Omega* **2021**, *6*, 9989−9997.

(15) Razzak, I.; Saris, R. A.; Blumenstein, M.; Xu, G. Integrating joint feature selection into subspace learning: A formulation of 2DPCA for outliers robust feature selection. *Neural Networks* **2020**, *121*, 441−451.

(16) Taqvi, S. A. A.; Zabiri, H.; Tufa, L. D.; Uddin, F.; Fatima, S. A.; Maulud, A. S. A review on data-driven learning approaches for fault detection and diagnosis in chemical processes. *ChemBioEng Rev.* **2021**, *8* (3), 239−259.

(17) Chen, L. F.; Liao, H. Y. M.; Ko, M. T.; Lin, J. C.; Yu, G. J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.* **2000**, *33* (10), 1713−1726.

(18) Ye, J.; Xiong, T. In *Null Space Versus Orthogonal Linear Discriminant Analysis*; Proceedings of the 23rd International Conference on Machine Learning ICML, 2006; pp 1073−1080.

(19) Wang, H.; Wang, H.; Nie, F.; Li, X. Ratio sum versus sum ratio for linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44* (12), 10171−10185.

(20) Zhu, F.; Gao, J.; Yang, J.; Ye, N. Neighborhood linear discriminant analysis. *Pattern Recog.* **2022**, *123*, No. 108422.

(21) Bertsimas, D.; Cory-Wright, R.; Pauphilet, J. Solving large-scale sparse PCA to certifiable (near) optimality. *J. Mach. Learn. Res.* **2022**, *23* (13), 1−35.

(22) Park, J.; Ahn, J.; Jeon, Y. Sparse functional linear discriminant analysis. *Biometrika* **2022**, *109* (1), 209−226.

(23) Hu, X.; Sun, Y.; Gao, J.; Hu, Y.; Ju, F.; Yin, B. Probabilistic linear discriminant analysis based on L 1-norm and its Bayesian variational inference. *IEEE Trans. Cybern.* **2022**, *52* (3), 1616−1627.

(24) Liang, Z.; Zhang, L. L_1-norm discriminant analysis via Bhattacharyya error bounds under Laplace distributions. *Pattern Recog.* **2023**, *141*, No. 109609.

(25) Li, R.; Wang, X.; Quan, W.; Song, Y.; Lei, L. Robust and structural sparsity auto-encoder with L_{21}-norm minimization. *Neurocomputing* **2021**, *425*, 71−81.

(26) Sha, X.; Diao, N. Robust kernel principal component analysis and its application in blockage detection at the turn of conveyor belt. *Measurement* **2023**, *206*, No. 112283.

(27) Nie, F.; Tian, L.; Huang, H.; Ding, C. Non-greedy l_{21}-norm maximization for principal component analysis. *IEEE Trans. Image Process.* **2021**, *30*, 5277−5286.

(28) Zhao, H.; Wang, Z.; Nie, F. A new formulation of linear discriminant analysis for robust dimensionality reduction. *IEEE Trans. Knowl. Data Eng.* **2019**, *31* (4), 629−640.

(29) Bounoua, W.; Bakdi, A. Fault detection and diagnosis of nonlinear dynamical processes through correlation dimension and fractal analysis based dynamic kernel PCA. *Chem. Eng. Sci.* **2021**, *229*, No. 116099.

(30) Ning, Z.; Xiao, Q.; Feng, Q.; Chen, W.; Zhang, Y. Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis. *IEEE Trans. Med. Imaging* **2021**, *40* (6), 1632−1645.

(31) Fu, Y.; Luo, C.; Bi, Z. Low-rank joint embedding and its application for robust process monitoring. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1−13.

(32) Odiowei, P. E. P.; Cao, Y. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Trans. Indus. Inform.* **2010**, *6*, 36−45.

(33) Sha, X.; Diao, N. l_{21}-Norm-Based robust feature extraction method for fault detection. *ACS Omega* **2022**, *7* (48), 43440−43449.