Locality Kernel Canonical Variate Analysis for Fault Detection

Shujun Xiao

Beijing University of Chemical Technology, Beijing 100029, China

E-mail: 15117955800@126.com

Abstract. Most of the existing fault detection techniques merely consider the global structure information, but neglect the local structure details between different data points. However, local structure information plays an essential role in data mining and feature extraction. In this article, a locality kernel canonical variate analysis (LPP-KCVA) method for fault detection is proposed. This method inherits the merits of KCVA in coping with nonlinear as well as dynamic problems, and makes full use of the advantage of Locality Preserving Projections (LPP) in retaining local manifold structure. To demonstrate the accuracy of the LPP-KCVA approach, Tennessee Eastman process (TEP) is applied, and KCVA, KPCA methods serve as the reference. The outcome illustrates its power in fault detection.

1. Introduction

With the speedy development of the modern industry and the increasing complexity in control systems, the capability to ensure product quality as well as the safe operation has considerably enhanced. Once a fault occurs, it will exert a detrimental impact on the stability of process, which can result in significant economic losses even heavy casualties. Over the past decades of years, fault detection has not only received gradually increasing concern, but an ocean of achievements have emerged as well. Basically, rough classification of fault detection methods is knowledge-based, model-based and data-driven ones [1]. Among the massive data-driven methods, multivariate statistical process monitoring (MSPM) ones are celebrated by the means of setting up thresholds obtained from mining inner features of historical data to distinguish whether a fault has occurred. Well-known methods of MSPM are those known as principal component analysis (PCA) [2, 3], partial least squares (PLS) [4, 5] and canonical variate analysis (CVA) [6-8].

As tools of dimensionality reduction, PCA and PLS are effective under the assumption of time independence and Gaussian distribution in static process [9]. However, due to the influence of noise and disturbance, actual process contains serial correlations which severely weaken their ability of capturing abnormalities. To address the drawbacks, variants of PCA and PLS [10, 11] consider the serial correlations between variables, but it is still challenging to manage strong autocorrelation in some circumstances. Compared with PCA and PLS, CVA could handle dynamic issue more proficient thanks to its consideration of both input and output data. It can’t be denied that CVA indeed possesses a wealth of excellence; however, it still bears some deficiency in most practical processes which exhibit nonlinear relationships between variables.

In an attempt to improve the insufficiency and uplift the performance of CVA, the kernel trick is integrated into CVA because it masters at disposing of nonlinear problems. The kernel method is popular for the fact that it does not call for settling down nonlinear optimization problem when
extracting nonlinear features [12]. As a novel method, kernel canonical variate analysis (KCVA) combines CVA with the kernel method to absorb their merits in accounting for dynamics and fully describing the nonlinear relationships between variables. On the basis of KCVA, some research productions have been conducted around the fault detection in recent years [13-16].

All the methods mentioned above merely focus on the information that the global structure holds, however, they ignore the local relationship between the data points. For the sake of inner data information and feature extraction, local structure methods have been presented to preferably interpret the deep links hidden behind the data. As an effective method for local information preservation, LPP can preserve the nonlinear structure of high-dimensional data after projection because of its ingenious combination of Laplacian Eigenmaps. Recently, combinations of MSPM methods and LPP have been extensively researched. Deng and Tian [17] extended LPP to KPCA for the sake of fault diagnosis and demonstrated its superiority over KPCA. Sun and Chen [18] proposed a LPCCA method which combines LPP with Canonical correlation analysis (CCA) for data visualization and pose estimation. Besides, Lu and Jiang [19] introduced a method for fault diagnosis based on the incorporation of CVA, FDA and LPP.

In this article, a method called locality kernel canonical variate analysis is suggested so as to monitor the process behaviour by taking both local and non-local structure information into consideration. The novel scheme is established on the basis of KCVA and LPP, which could handle nonlinearity, serial correlations issues and could retain local manifold structure simultaneously. The article can be separated into the following parts. A brief review of CVA, KCVA and LPP will be given in Section2. Section 3 will present a detailed introduction on LPP-KCVA. To illustrate the effectiveness of LPP-KCVA, Section 4 is used to focus on the simulation of the TEP. Finally, some conclusions are drawn in Section 5.

2. CVA, KCVA and LPP Method

2.1. CVA algorithm

As a subspace identification method, the linear state-space model of CVA is expressed as

\[ x_{t+1} = \Phi x_t + Gu_t + w_t \]  \[ y_t = Hx_t + Au_t + BW_t + v_t \]

where \( u_t \in \mathbb{R}^{m_u} \) is the input vector, \( x_t \in \mathbb{R}^d \) is the state vector, \( y_t \in \mathbb{R}^{m_y} \) is the output vector, \( w_t \) and \( v_t \) are independent stochastic white noise with zero mean, and \( \Phi, G, H, A, B \) are state space matrices. Define \( p_t \) and \( f_t \) as the past and future vectors at time point \( t \), then they can be stacked as:

\[ p_t = [y_{t-1}^T, y_{t-2}^T, \ldots, y_{t-l}^T, u_{t-1}^T, u_{t-2}^T, \ldots, u_{t-l}^T]^T \]

\[ f_t = [y_{t+h}^T, y_{t+h+1}^T, \ldots, y_{t+h+l}^T]^T \]

where \( l \) and \( h \) denote the lag number chosen to better address autocorrelation issues. Both \( p_t \) and \( f_t \) vectors are normalized to eliminate the dimension effect among different observation variables. If \( s \) observations are available, then the Hankel matrices for \( p_t \) and \( f_t \) are collected in the following form:

\[ X_p = [p_t, p_{t+1}, \ldots, p_{t+n-1}] \]

\[ X_f = [f_t, f_{t+1}, \ldots, f_{t+n-1}] \]

where \( X_p \in \mathbb{R}^{(m_u+m_y)\times N}, X_f \in \mathbb{R}^{(m_y+l)\times N} \), and \( N \) is the column number in Hankel matrices, \( N = s - l - h + 1 \).

Then it is naturally to obtain the auto covariance matrices of \( X_p, X_f \) as well as their cross-covariance. \( \Sigma_{pp}, \Sigma_{ff}, \Sigma_{pf} \) denote the auto covariance matrices of \( X_p, X_f \) and their cross-covariance respectively. Singular value decomposition can provide a solution for the CVA method.

\[ \Sigma_{pp}^{-1/2} \Sigma_{pf} \Sigma_{ff}^{-1/2} = U \Sigma V^T \]

For ease of calculation, only the first \( d \) (< \( r \)) canonical variables are selected which could account for most of the dynamic properties. Therefore, the state vector resembles the canonical variable can be calculated as:
\[ x_d(t) = J_d p_t = U_d^T \Sigma_{d,p}^{-1/2} p_t \]  
(8)

where \( J_d = U_d^T \Sigma_{d,p}^{-1/2} \). \( U_d \) represents the column vectors corresponding to the first \( d \) maximum correlation coefficients in \( U \).

### 2.2. KCVA algorithm

Generally, fewer studies have expanded the CVA method to the kernel field. The kernel method could transform the linear inseparable data in low dimension into the linear separable data in high dimension by suitable non-linear mapping functions. On combining CVA with the kernel method, KCVA provides relative ease to analyze large-scale and complex industrial process. Similarly, let \( \varphi_1(X_p) \) and \( \varphi_2(X_f) \) denote the matrices projected from the kernel matrices, and then the KCVA problem can be transformed into the following form:

\[
\begin{bmatrix}
0 & \varphi_1(X_p) \varphi_2^T(X_f) \\
\varphi_2(X_f) \varphi_1^T(X_p) & 0
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
= \lambda
\begin{bmatrix}
\varphi_1(X_p) \varphi_1^T(X_p) & 0 \\
0 & \varphi_2(X_f) \varphi_2^T(X_f)
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\]  
(9)

\( \tilde{\alpha} \) can be regarded as the vector spanned by the columns of \( \varphi_1(X_p) \), \( \tilde{\alpha} = \varphi_1(X_p) \alpha \). Similarly, \( \tilde{\beta} = \varphi_2(X_f) \beta \).

Now, using the kernel trick like:

\[
\begin{bmatrix}
K_{pp} & K_{pf} \\
K_{fp} & K_{ff}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
= \lambda
\begin{bmatrix}
\tilde{K}_{pp} & \tilde{K}_{pf} \\
\tilde{K}_{fp} & \tilde{K}_{ff}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\]  
(10)

where

\[
\begin{bmatrix}
K_{pp}
\end{bmatrix}_{ij} = k_p(X_{p,i}, X_{p,j}) = \langle \varphi_1(X_{p,i}), \varphi_1(X_{p,j}) \rangle
\]  
(12)

\[
\begin{bmatrix}
K_{pf}
\end{bmatrix}_{ij} = k_f(X_{f,i}, X_{f,j}) = \langle \varphi_2(X_{f,i}), \varphi_2(X_{f,j}) \rangle
\]  
(13)

\( X_{p,i} \) is the \( i \)th column in \( X_p \) and \( X_{f,i} \) is the \( i \)th column in \( X_f \), both \( K_{pp} \) and \( K_{ff} \) are Gram matrices which centered as \( \bar{K}_{pp} = K_{pp} - K_{pp} E - E K_{pp} + EK_{pp}E \) and \( \bar{K}_{ff} = K_{ff} - K_{ff} E - E K_{ff} + EK_{ff}E \). The kernel functions \( k_p, k_f \) usually choose radial basis function. By applying kernel function, the equation (9) can be rewritten as:

\[
\begin{bmatrix}
0 & \bar{K}_{pp} \bar{K}_{ff} \\
\bar{K}_{ff} \bar{K}_{pp} & 0
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
= \lambda
\begin{bmatrix}
\bar{K}_{pp} & 0 \\
0 & \bar{K}_{ff}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\]  
(14)

In order to tackle the above optimization problem, generalized eigenvalue decomposition can be applied. One thorny problem is that \( \bar{K}_{pp} \) and \( \bar{K}_{ff} \) are singular, thus \( \bar{K}_{pp} \bar{K}_{pp} + \eta I_1 \) and \( \bar{K}_{ff} \bar{K}_{ff} + \eta I_2 \) serve as the replacement of \( \bar{K}_{pp} \bar{K}_{pp} \) and \( \bar{K}_{ff} \bar{K}_{ff} \). The matrix functions \( \eta_1 \) and \( \eta_2 \) serve as the replacement of \( \bar{K}_{pp} \bar{K}_{pp} \) and \( \bar{K}_{ff} \bar{K}_{ff} \). The kernel functions \( k_p, k_f \) usually choose radial basis function. By applying kernel function, the equation (9) can be rewritten as:

\[
\begin{bmatrix}
0 & \bar{K}_{pp} \bar{K}_{ff} \\
\bar{K}_{ff} \bar{K}_{pp} & 0
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
= \lambda
\begin{bmatrix}
\bar{K}_{pp} & 0 \\
0 & \bar{K}_{ff}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\]  
(14)

The coefficient vectors \( \alpha \) and \( \beta \) can be obtained by computing equation (14). For the past vector \( p_t \), the canonical vector \( c \) can be calculated as:

\[
c = A^T \varphi_1(p_t) = A^T \varphi_1^T(X_p) \varphi_1(p_t) = A^T k_p(X_p, p_t)
\]  
(15)

where \( A = [\alpha_1, \alpha_2, \cdots, \alpha_N] \), \( A = [\alpha_1, \alpha_2, \cdots, \alpha_N] \), and \( k_p(X_p, p_t) = [k_p(X_{p_1}, p_t), k_p(X_{p_2}, p_t), \cdots, k_p(X_{p_N}, p_t)]^T \).

### 2.3. LPP algorithm

LPP is a method aiming to capture the local manifold structures among samples in the initial sample space and maintains their previous relationships after the projection. More specifically, if two data points are neighbors in the high-dimensional data space, they are close as well after projected to the low-dimensional data space. The purpose of LPP is to avoid the divergence of data points and maintain the original local structure.

Given the original data sample matrix \( X = [x_1, x_2, \cdots, x_n] \), where \( n \) is the number of samples. And define \( g \) as the projection vector that transforms \( X \) to \( Z \), where \( Z = [z_1, z_2, \cdots, z_n] \). \( z_i = g^T x_i \), \( i = 1, \cdots, n \). The objective function of LPP algorithm is expressed in the following form:
where \( w_{ij} \) is a weighting parameter that represents the relationship between the two samples and it can be calculated by the heat kernel function:

\[
\exp \left( -\frac{||x_i - x_j||^2}{\sigma} \right), \text{if } x_i \in \text{Neigh}(x_j) \text{ or } x_j \in \text{Neigh}(x_i)
\]

otherwise

where \( \text{Neigh}(x_j) \) denotes the neighbors of \( x_j \) and \( k \)-Nearest Neighbor method can be applied to judge the range of the neighbors. It can be seen that, for each data point, a non-zero weight is given to the data point next to it, and a zero value is given to the data point far away from it, then the goal of preserving the local structure of the data can be achieved. Formula (16) can be further deduced as follows:

\[
J = \min \sum_{i=1}^{n} \sum_{j=1}^{n} (z_i - z_j)^2 w_{ij} = \min \sum_{i=1}^{n} \sum_{j=1}^{n} g^T(x_i - x_j) w_{ij}(x_i - x_j)^T g
\]

where \( w_{ij} \) is a weighting parameter that represents the relationship between the two samples and it can be calculated by the heat kernel function:

\[
\exp \left( -\frac{||x_i - x_j||^2}{\sigma} \right), \text{if } x_i \in \text{Neigh}(x_j) \text{ or } x_j \in \text{Neigh}(x_i)
\]

otherwise

where \( \text{Neigh}(x_j) \) denotes the neighbors of \( x_j \) and \( k \)-Nearest Neighbor method can be applied to judge the range of the neighbors. It can be seen that, for each data point, a non-zero weight is given to the data point next to it, and a zero value is given to the data point far away from it, then the goal of preserving the local structure of the data can be achieved. Formula (16) can be further deduced as follows:

\[
J = \min \sum_{i=1}^{n} \sum_{j=1}^{n} g^T(x_i - x_j) w_{ij}(x_i - x_j)^T g = \min \sum_{i=1}^{n} \sum_{j=1}^{n} g^T(x_i - x_j) w_{ij}(x_i - x_j)^T g
\]

where \( L_x = D_x - W \), \( W \) stands for the weighting matrix, \( L_x \) is the Laplace matrix and \( D = \text{diag}(d_1, d_2, \cdots, d_n) \) with \( d_i = \sum_{j=1}^{n} w_{ij} \). In order to prevent over-fitting, some restrictions need to be added as follows:

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} g^T(x_i - x_j) w_{ij}(x_i - x_j)^T g = 1
\]

then equation (18) can be written as:

\[
g^T X D X^T g = 1
\]

That is to say, under the condition of \( g^T X D X^T g = 1 \), the value of \( g^T X L_x X^T g \) should be minimized.

\[
\min g^T X L_x X^T g
\]

\[
s.t. g^T X D X^T g = 1
\]

In this paper, the main function of LPP is to retain the local structure and avoid losing the details.

3. Locality Kernel Canonical Variate Analysis (LPP-KCVA) method

After projected from the original low-dimensional space to high-dimensional space, the data may not be able to maintain the original local structure information due to the data divergence. In this section, a fault detection method called locality kernel canonical variate analysis (LPP-KCVA) is proposed to solve the above problem. With some constraints provided by LPP, the data could retain local manifold in low-dimensional space and avoid the loss of detailed information even if the dimension changes. In addition, LPP-KCVA absorbs the merits of LPP and KCVA, so it is proficient in addressing the mixed problems.

As defined in equation (5) and (6), \( X_p \) and \( X_f \) are the Hankel matrices respectively, \( \phi_1(X_p) \) and \( \phi_2(X_f) \) denote the matrices projected from \( X_p \) and \( X_f \). Consider the LPP algorithm and the KCVA method introduced in section 2, if \( X \) in equation (21) is replaced by \( \phi_1(X_p) \), then we can obtain the following equation:

\[
J_p = \min g^T \phi_1(X_p) L_p \phi_1^T(X_p) g
\]

where \( L_p \) denotes the Laplace matrix for the past data. Similar to the KCVA method, \( g \) can be spanned by the column vectors of \( \phi_1(X_p) \).

\[
g = \phi_1(X_p) \alpha
\]

By replacing \( g \) in equation (22), the objective function \( J_p \) can be written as:

\[
J_p = \min \{ \alpha^T \phi_1^T(X_p) L_p \phi_1(X_p) \phi_1^T(X_p) \phi_1(X_p) \alpha \}
\]

After implementing standardization on \( K_{pp} \), equation (24) can be rewritten as:

\[
J_p = \min \alpha^T \tilde{K}_{pp} L_p \tilde{K}_{pp} \alpha
\]

Similarly, the optimal function of Hankel matrix \( X_f \) can be obtained:

\[
J_f = \min \beta^T \tilde{K}_{pf} L_f \tilde{K}_{pf} \beta
\]

On combining with the conditions of KCVA, the following optimization problem can be obtained:
The above optimization problem can be transformed into the following equation:

\[
\max \alpha^T K_{pp} \beta = 1, \beta^T R_{ff} L_{r} R_{ff} = 1
\]

The above equation is of the form (27) and the corresponding optimization problem is of the form (28) when the generalized eigenvalue decomposition can solve the above equation. Like the KCVA method, because \( R_{pp} \) and \( R_{ff} \) are singular matrices, \( R_{pp}^+ \) and \( R_{ff}^+ \) should be used rather than \( R_{pp} \) and \( R_{ff} \). Then the canonical variable \( c \) of the LPP-KCVA method can be obtained:

\[
c = A^T \phi_1(p_t) = A^T \phi_1^T(X_p) \phi_1(p_t) = A^T K_p (X_p, p_t)
\]

where

\[
A = [\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_N], k_p(X_{p}, p_t) = [k_p(X_{p,1}, p_t), k_p(X_{p,2}, p_t), \cdots, k_p(X_{p,N}, p_t)]^T.
\]

Constructing statistics \( T_s^2, T_r^2 \) corresponding to the state space and the residual space

\[
T_s^2 = c_s^T c_s
\]

\[
T_r^2 = c_r^T c_r
\]

where \( c_s \) comprises the first s column of \( c \) while \( c_r \) comprises the rest of the column of \( c \).

Accordingly, the control limits \( T_{s, UCL}^2 \), \( T_{r, UCL}^2 \) of \( T_s^2 \) and \( T_r^2 \) statistics can be calculated as below:

\[
\int_{-\infty}^{+\infty} p(T_s^2) dT_s^2 = \gamma \quad (32)
\]

\[
\int_{-\infty}^{+\infty} p(T_r^2) dT_r^2 = \gamma \quad (33)
\]

The control limits \( T_{s, UCL}^2 \) and \( T_{r, UCL}^2 \) can be estimated provided that \( p(T_s^2), p(T_r^2) \) and the significance level \( \gamma \) are determined beforehand. Through comparing the value of \( T_s^2 \) and \( T_r^2 \) and their control limits, a fault can be captured if the value of \( T_s^2 \) is larger than \( T_{s, UCL}^2 \) or the value of \( T_r^2 \) is larger than \( T_{r, UCL}^2 \).

To sum up, the LPP-KCVA procedure can be briefly stated as below:

Off-line model establishment

1. Prepare the data under normal condition and normalize each variable.
2. Define the past and future vectors by using equation (3), (4) and put them into the past and future Hankel matrices \( X_p \) and \( X_f \) like equation (5), (6).
3. Select proper kernel functions \( k_p, k_f \) and calculate the Gram matrices \( K_{pp}, K_{ff} \) with equation (10), (11).
4. Implement the normalization on the matrices \( K_{pp}, K_{ff} \).
5. Calculate the weighting matrices \( W_p, W_f \), diagonal matrices \( D_p, D_f \) and Laplace matrices \( L_p \) and \( L_f \).
6. Solve the generalized eigenvalue problem in equation (28) to obtain \( \alpha \) and \( \beta \).
7. Compute the \( T_s^2 \) and \( T_r^2 \) statistics with equation (30), (31) and calculate the control limits \( T_{s, UCL}^2 \) and \( T_{r, UCL}^2 \) using equation (32), (33).

Online monitoring

1. Scale the new test data with the mean and variance identical to the training data.
2. Form the online \( p_k \) and \( f_k \) vectors and calculate each k row of Gram matrices \( K_{pp}^{test} \) and \( K_{ff}^{test} \) with \[ K_{pp}^{test} \] and \( K_{ff}^{test} \) as follows:

\[
K_{pp}^{test} = k_p^T (X_p, f_k) \quad \text{and} \quad K_{ff}^{test} = k_f^T (X_f, f_k).
\]
3. Center the Gram matrices \( K_{pp}^{test} \) and \( K_{ff}^{test} \) as follows:

\[
K_{pp}^{test} = K_{pp}^{test} - \frac{1}{N} k_p (X_p, f_k) E - k_p (X_p, f_k) E - 1^{test} K_{pp}^{test} + 1 \quad \text{and} \quad K_{ff}^{test} = K_{ff}^{test} - \frac{1}{N} k_f (X_f, f_k) E - k_f (X_f, f_k) E - 1^{test} K_{ff}^{test} + 1 \quad \text{where} \quad 1^{test} = 1/N [1, \cdots, 1] \in \mathbb{R}^{N \times N}.
\]
4. Calculate \( T_{s,k}^2 \) and \( T_{r,k}^2 \) at each time point k with equation (30), (31).
5. Compare the obtained \( T_{s,k}^2 \) and \( T_{r,k}^2 \) with \( T_{s,UCL}^2 \) and \( T_{r,UCL}^2 \) to decide whether a fault has occurred.

When a fault exists, \( T_{s,k}^2 > T_{s,UCL}^2 \) or \( T_{r,k}^2 > T_{r,UCL}^2 \).
4. Experiments on the Tennessee Eastman Process
In this section, the Tennessee Eastman process (TEP) is utilized to assess the effectiveness of the proposed LPP-KCVA fault detection technique. The TEP establishes a realistic nonlinear dynamic industrial platform for assessing newly developed fault detection methods. It has five major reactions and eight components. The process covers 21 faults, one normal condition and involves 52 variables in which 12 are manipulated variables and 41 are measurement variables. The training data include 500 samples under the fault-free condition and 480 samples under the faulty conditions. The testing data include 960 samples for each fault. Each fault begins at the 161th sample and each data is sampled every 3 minutes.

The performance of the developed LPP-KCVA method is assessed on the foundation of the missed detection rate which is based on the percentage of the data sample value smaller than the control limits in the last 800 faulty data sample in 21 faults. The missed detection rate reveals the sensitivity of each fault method, more precisely, if the missed detection rate of one method is lower than others, then the method can be considered better. Another index for monitoring the performance of the LPP-KCVA method is detection delay which measures the promptness. The detection delay is referred as the time interval which the fault introduced to the system until the fault to be detected. From the perspective of the promptness, if one fault detection method has shorter detection delay time, it can be considered better than others.

As the false alarms are inescapable, it might be considered faulty under the fault-free conditions. To reduce the false alarms, a fault can be identified if several continuous statistics have the value smaller than the control limits.

In Table 1 displays the missed detection rates for the testing set.

| Fault | KCVA $T^2_2$ | KCVA $T^2_2$ | KPCA $T^2$ | KPCA $Q$ | LPP-KCVA $T^2_2$ | LPP-KCVA $T^2_2$ |
|-------|--------------|--------------|------------|----------|----------------|----------------|
| 1     | 0.0038       | 0.0013       | 0.0063     | 0        | 0.0025         | 0              |
| 2     | 0.0165       | 0.0114       | 0.0138     | 0.0063   | 0.0127         | **0.0025**     |
| 3     | 0.6810       | 0.9899       | 0.9437     | 0.7612   | 0.8241         | **0.3304**     |
| 4     | 0.8747       | 0.8430       | 0.3250     | 0        | 0.9304         | 0              |
| 5     | 0.5911       | 0.5557       | 0.7212     | 0.4163   | 0.6329         | **0.3203**     |
| 6     | 0.0013       | 0            | 0.8962     | 0        | 0.0013         | 0              |
| 7     | 0.4278       | 0.4354       | 0.0213     | 0        | 0.1152         | 0              |
| 8     | 0.0089       | 0.0481       | 0.0262     | 0.0088   | 0.0114         | **0.0013**     |
| 9     | 0.7203       | 0.9924       | 0.9425     | 0.7963   | 0.7456         | **0.3456**     |
| 10    | 0.3430       | 0.3228       | 0.5225     | 0.1737   | 0.4228         | 0.2481         |
| 11    | 0.6962       | 0.7114       | 0.4150     | 0.1475   | 0.7215         | **0.0139**     |
| 12    | 0.0177       | 0.1038       | 0.0250     | 0.0025   | 0.0152         | 0              |
| 13    | 0.0418       | 0.0418       | 0.0462     | 0.0413   | 0.0367         | **0.0253**     |
| 14    | 0.9797       | 0.9886       | 0          | 0        | 0.0038         | 0              |
| 15    | 0.7456       | 0.6291       | 0.9263     | 0.7025   | 0.7899         | **0.4671**     |
| 16    | 0.3937       | 0.8810       | 0.6987     | **0.2000** | 0.5127         | 0.2468         |
| 17    | 0.2873       | 0.2937       | 0.2338     | 0.0213   | 0.1570         | **0.0127**     |
| 18    | 0.0835       | 0.0696       | 0.8375     | 0.0737   | 0.0924         | 0.0215         |
| 19    | 0.8949       | 0.9684       | 0.8712     | 0.3413   | 0.9316         | **0.0975**     |
| 20    | 0.4291       | 0.4354       | 0.5300     | 0.2100   | 0.4342         | 0.0595         |
| 21    | 0.6177       | 0.7101       | 0.5550     | 0.3175   | 0.6165         | **0.1367**     |

Table 1 displays the missed detection rates about the $T^2_2$, $T^2_2$ statistics in LPP-KCVA, the $T^2_2$, $T^2_2$ statistics in KCVA and the $T^2$ and $Q$ statistics in KPCA for 21 faults. The bold figures represent the best statistic in one fault. It can be seen from the table that, among all the 21 faults, the $T^2$ statistic in
LPP-KCVA acts superior to other statistics in terms of missed detection rates. The $T^2$ statistic in LPP-KCVA possesses the best performance in most fault conditions. However, the $T^2$ statistic still has some drawbacks when facing the Fault 3 and Fault 9. The superiority of the presented method tends to owe to the consideration of both local and global information.

Table 2. Detection delay (minutes) for the testing set

| Fault | KCVA $T^2$ | KCVA $T^2$ | KPCA $T^2$ | KPCA $Q$ | KPCA $T^2$ | LPP-KCVA $T^2$ | LPP-KCVA $T^2$ |
|-------|------------|------------|------------|----------|------------|----------------|----------------|
| 1     | 9          | 3          | 12         | 0        | 6          | 0              | 0              |
| 2     | 39         | 27         | 33         | 0        | 30         | 6              | 6              |
| 3     | 3          | 1188       | 60         | 24       | 51         | 3              | 3              |
| 4     | 162        | 159        | 0          | 0        | 171        | 0              | 0              |
| 5     | 27         | 0          | 3          | 3        | 30         | 0              | 0              |
| 6     | 3          | 0          | 15         | 0        | 3          | 0              | 0              |
| 7     | 0          | 0          | 0          | 0        | 0          | 0              | 0              |
| 8     | 0          | 15         | 45         | 18       | 12         | 3              | 3              |
| 9     | 0          | 984        | 0          | 0        | 3          | 0              | 0              |
| 10    | 12         | 12         | 21         | 0        | 18         | 9              | 9              |
| 11    | 96         | 366        | 15         | 3        | 27         | 3              | 3              |
| 12    | 12         | 69         | 6          | 6        | 12         | 0              | 0              |
| 13    | 0          | 0          | 108        | 51       | 87         | 3              | 3              |
| 14    | 90         | 2112       | 0          | 0        | 0          | 0              | 0              |
| 15    | 261        | 30         | 273        | 168      | 1413       | 30             | 30             |
| 16    | 0          | 906        | 12         | 24       | 0          | 3              | 3              |
| 17    | 96         | 93         | 3          | 21       | 78         | 3              | 3              |
| 18    | 153        | 150        | 51         | 30       | 42         | 0              | 0              |
| 19    | 0          | 201        | 30         | 3        | 18         | 0              | 0              |
| 20    | 216        | 219        | 0          | 0        | 171        | 3              | 3              |
| 21    | 1236       | 1683       | 39         | 15       | 930        | 249            | 249            |

The details of the detection delay for the LPP-KVCA, KCVA and KPCA methods are tabulated in Table 2. As the Table 2 presents, the $T^2$ statistic in LPP-KCVA can capture the fault earlier than other statistics for most faults. In other words, the $T^2$ statistic in LPP-KCVA can contribute more time for operators to deal with the abnormalities.

Take Fault 7 as an example to implement fault detection. The real-time monitoring curves are displayed in Figure 1-6. For the KPCA method, the $T^2$ statistic appears missed detection for several times after the occurrence of Fault 7 while the Q statistic appears several false alarms in the first 160 samples. The $T^2$ and $T^2$ statistics of KCVA possess relatively high missed detection rates though no false alarms appear under normal conditions. The $T^2$ statistic in LPP-KCVA method can basically realize the function of fault detection for Fault 7. As for the $T^2$ statistic in LPP-KCVA method, it has the lowest missed detection rate among all the statistics and bears no false alarms. Thus, the $T^2$ statistic of the LPP-KCVA method can detect faults more accurately and timely.
8

5. Conclusion
This paper proposed a novel fault detection method called the locality kernel canonical variate analysis (LPP-KCVA), which integrates the ability of KCVA in handling the nonlinearity and serial
correlations with the capacity of LPP in preserving the local manifold structure. The LPP-KCVA method takes both local and non-local structure information into consideration to get more details. The efficiency of this method is verified by implementing experiments on Tennessee Eastman process in comparison with KCVA and KPCA. The outcome shows that the LPP-KCVA scheme presents a superior performance for fault detection than KCVA and KPCA.

References
[1] Chiang L H, Russell E L, Braatz R D. Fault detection and diagnosis in industrial systems[M]. Springer Science & Business Media, 2001.
[2] Nomikos P, MacGregor J F. Monitoring batch processes using multiway principal component analysis[J]. AIChE Journal, 1994, 40(8): 1361-1375.
[3] Bakshi B R. Multiscale PCA with application to multivariate statistical process monitoring[J]. AIChE journal, 1998, 44(7): 1596-1610.
[4] Ji H, He X, Shang J, et al. Incipient fault detection with smoothing techniques in statistical process monitoring[J]. Control Engineering Practice, 2017, 62: 11-21.
[5] Zhu Q, Liu Q, Qin S J. Concurrent canonical correlation analysis modeling for quality-relevant monitoring[J]. IFAC-PapersOnLine, 2016, 49(7): 1044-1049.
[6] Russell E L, Chiang L H, Braatz R D. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis[J]. Chemometrics and Intelligent laboratory systems, 2000, 51(1): 81-93.
[7] Juricek B C, Seborg D E, Larimore W E. Fault detection using canonical variate analysis[J]. Industrial & engineering chemistry research, 2004, 43(2): 458-474.
[8] Jiang B, Braatz R D. Fault detection of process correlation structure using canonical variate analysis-based correlation features[J]. Journal of Process Control, 2017, 58: 131-138.
[9] Joe Qin S. Statistical process monitoring: basics and beyond[J]. Journal of the Chemometrics Society, 2003, 17(8-9): 480-502.
[10] Ku W, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis[J]. Chemometrics and intelligent laboratory systems, 1995, 30(1): 179-196.
[11] Lee G, Song S O, Yoon E S. Multiple-fault diagnosis based on system decomposition and dynamic PLS[J]. Industrial & engineering chemistry research, 2003, 42(24): 6145-6154.
[12] Cho J H, Lee J M, Choi S W, et al. Fault identification for process monitoring using kernel principal component analysis[J]. Chemical engineering science, 2005, 60(1): 279-288.
[13] Odiowei P E P, Cao Y. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations[J]. IEEE Transactions on Industrial Informatics, 2010, 6(1): 36-45.
[14] Giantomassi A, Ferracuti F, Iarlori S, Longhi S, Fonti A, Comodi G. Kernel canonical variate analysis based management system for monitoring and diagnosing smart homes[C]. International Joint Conference on Neural Networks. IEEE, 2014: 1432-1439.
[15] Samuel R T, Cao Y. Improved kernel canonical variate analysis for process monitoring[C]. Automation and Computing. IEEE, 2015: 1-6.
[16] Linzhe H, Yuping C, Xuemin T, et al. A nonlinear quality-relevant process monitoring method with kernel input-output canonical variate analysis[J]. IFAC-PapersOnLine, 2015, 48(8): 611-616.
[17] Deng X, Tian X, Chen S. Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis[J]. Chemometrics and Intelligent Laboratory Systems, 2013, 127: 195-209.
[18] Sun T, Chen S. Locality preserving CCA with applications to data visualization and pose estimation[J]. Image and Vision Computing, 2007, 25(5): 531-543.
[19] Lu Q, Jiang B, Gopaluni R B, Loewen P D, Braatz R D. Locality preserving discriminative canonical variate analysis for fault diagnosis[J]. Computers & Chemical Engineering, 2018, 117: 309-319.