Process monitoring based on generalized orthogonal neighborhood preserving embedding


State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, Zhejiang, China
(e-mail: am_miao@zju.edu.cn, songzihuan@zju.edu.cn, wenqiaojun@zju.edu.cn, gezhiqiang@zju.edu.cn)

Abstract: A new orthogonal neighborhood preserving embedding (ONPE) and its kernel generalization based process monitoring approaches are presented in this paper. ONPE aims at preserving local neighborhood structure of process data while reducing data dimensionality. As an approximation of the nonlinear manifold learning method, ONPE is capable to handle process nonlinearity. Moreover, to enhance the nonlinear modeling performance, the nonlinear extension of ONPE is also developed, with the introduction of kernel-tricks. By constructing monitoring statistics, both ONPE and its generalization are applied for fault detection in nonlinear processes. Two case studies show the superiority of the proposed methods in process monitoring.

1. INTRODUCTION

Data based multivariate statistical techniques have become one of the research hotspots in industrial process monitoring. Principal component analysis (PCA) and partial least squares (PLS) are two representative techniques for extracting useful information and have been effectively applied for process monitoring. Compared to traditional methods, the data driven techniques are more sophisticated for addressing high dimensional, strongly correlated process data. As one of the key data characteristics, the nonlinearity has recently been brought into focus in this area. For example, several extensions of PCA and PLS have been extensively reported, each of which is characterized by focusing on specific process aspects in order to get better monitoring performance, e.g. Hiden (1999) combined genetic programming and linear PCA to produce a non-linear PCA algorithm; Maulud (2006) proposed an orthogonal nonlinear PCA based on the wavelet decomposition; Kamppjärvi(2008) introduced neural network and the PCA based fault detection and isolation system. Besides, the kernel methods have also been successfully utilized for nonlinear process monitoring (Lee et al., 2004; Shao et al., 2009). Without the tedious nonlinear optimization procedure, the kernel method provides a better understanding of the nonlinear structure features and the implementation is also quite easy. In the past years, several kernel based monitoring methods had been proposed, such as KPCA (Ge et al., 2009), kernel LDA (Zhu et al., 2011) and kernel ICA (Zhang et al., 2007).

Recently, several new nonlinear dimensionality reduction algorithms named manifold learning have been proposed, such as isometric mapping (ISOMAP) (Tenenbaum et al., 2000), locally linear embedding (LLE) (Roweis et al., 2000), Laplacian eigenmap (LE) (Belkin et al., 2002), local tangent space alignment (LTSA) (Zhang et al., 2003) and so on. In manifold learning, nonlinear dimensionality reduction is achieved based on certain properties of the dataset to be preserved. Like LLE, it constructs a local geometric structure and tries to preserve those local geometries in a low dimensional space through an optimal way. Although LLE is a good dimensionality reduction technique to deal with the nonlinear problems, it yields mappings that are defined only on the training data points, thus it remains unclear how to naturally evaluate the maps on the new testing data points. Neighborhood preserving embedding (NPE) is the linear approximation to LLE algorithm (He et al., 2005). The NPE inherits the ability of LLE in preserving the local manifold structure of the dataset; it is different from conventional PCA which is good at preserving the global Euclidean structure. NPE seeks to map close points on the original space to close in the low dimensional representation, while PCA maps the faraway points nearby (Hu et al., 2008). As it is derived by preserving neighbourhoud information of the data, NPE is less sensitive to noise compared to PCA. Due to its powerful data processing performance, the NPE method has been successfully applied in fed-batch process monitoring (Hu et al., 2008) and face recognition (Wang et al., 2010).

Based on the conventional NPE algorithm, an orthogonal form has also been developed, which has an additional orthogonal constraint on the basic functions (Liu et al., 2007). Therefore, it shares the same local preserving property with NPE and LLE. The orthogonal basis functions preserve the metric structure of the original data space, and avoid the metric distortion problem of NPE (Cai et al., 2006). The orthogonal property also has advantages in process monitoring. Firstly, orthogonal basis functions can successfully avoid the singularity problem in NPE, thus it owns superiority in statistics reconstruction and computing reconstruction error, based on which the fault detection performance can be improved. Secondly, the orthogonal feature can effectively improve the ability of discriminating, which is helpful to fault identification (Shao et al., 2009).

In this work, we introduce the ONPE method for nonlinear process monitoring. Since ONPE is essentially a linear dimensionality technique, its ability in dealing with the
nonlinearity of the process data is limited. Therefore, a kernel orthogonal neighborhood preserving embedding method (KONPE) is proposed for monitoring performance enhancement in nonlinear processes by means of the kernel trick. Similar to other kernel methods, the basic idea of the proposed KONPE method is to map the input space into a high dimension space via a nonlinear map, and then linear ONPE transform is performed in the feature space. This paper is organized as follows: First, the ONPE algorithm is briefly introduced in section 2. The proposed KONPE is described in section 3. Section 4 provides detailed fault detection method based on ONPE and KONPE. Section 5 gives a numerical example and the TE benchmark process simulation studies, which shows the effectiveness of the proposed methods. Conclusions are made in the last section.

2. ORTHOGONAL NEIGHBORHOOD PRESERVING EMBEDDING

Consider a dataset \(X(x_1,\ldots,x_n) \in \mathbb{R}^{m \times n}\) (measured variable \(m\), sampling points \(n\)) in the original high dimensional space. In the LLE method, an adjacency graph is built to preserve the local geometric structure of the data manifold and the relationship between the sample points. The local geometry is reflected by the linear coefficients that reconstruct each data point from its neighbours. Let \(W\) denote the weight matrix, where \(W_{ij}\) represents the linear combination coefficients to reconstruct \(x_i\) by using its nearest neighbour \(x_j\), the matrix \(W\) is computed by minimizing the following error function:

\[
\min \sum_j \left\| x_i - \sum_j W_{ij} x_j \right\|^2
\]

In the low dimensional subspace \(Y(y_1,\ldots,y_n) \in \mathbb{R}^d\) \((d < m)\), such local neighbourhood properties are preserved by retaining the same reconstruction weights between the point \(y_i\) and its corresponding neighbours. The optimal lower dimensional embedding \(Y\) are obtained by optimizing the formulation (2):

\[
\min \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2
\]

Under the constraint that \(\sum_j W_{ij} = 1\) \((i = 1,2,3,\ldots,n,\ j = 1,2,3,\ldots,k)\). If \(x_i\) is not the neighbour of \(x_j\), \(W_{ij} = 0\). \(Y\) is constrained to have unit covariance, that is \(Y^T Y = a^T X X^T a = I\).

Without a definite transformation matrix, LLE is defined only on the training data points and cannot handle the new data points (He et al., 2005). As a linear approximation algorithm to LLE, NPE shares the same neighbourhood preserving character as LLE. Meanwhile, in the NPE method, a linear transformation matrix \(a(a_1,\ldots,a_d) \in \mathbb{R}^{m \times d}\) is generated from the training data to relate the high dimension original data to the mapping data, that is: \(Y = a^T X\). Any new test data can be evaluated to locate in the reduced representation space by the matrix \(a\). In ONPE, an orthogonal constraint is imposed on the columns of the transformation matrix \(a(a_1,\ldots,a_d)\). The objective function of ONPE becomes:

\[
a_i = \arg\min_a \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 = \arg\min_a a^T X X^T a
\]

\[
\text{s.t. } a^T X X^T a = 1
\]

\[
a_i = \arg\min_a \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 = \arg\min_a a^T X X^T a
\]

\[
\text{s.t. } a_i^T a_j = a_j^T a_i = a_i^T a_{i+1} = 0
\]

\[
a^T X X^T a = 1
\]

Where \(k = 2,\ldots,d, M = (I - W)(I - W)^T\). Using the Lagrange multipliers, the above objective functions are transformed to include all the constraints. The orthogonal vectors \(a\) is finally calculated iteratively as follows:

1. \(a_i\) is computed as the eigenvector of \((X X^T)^{-1} X M X^T\) associated with the smallest eigenvalues.
2. \(a_i\) is computed as the eigenvectors associated with the smallest eigenvalues of \(Q^{(i)}\):

\[
Q^{(i)} = \left\{1 - (X X^T)^{-1} a_i (a_i)^T S^{(i-1)} (a_i)^T X X^T\right\} (X X^T)^{-1} X M X^T
\]

\[
S^{(i-1)} = \left( a_i (a_i)^T X X^T\right)^{-1} a_i (a_i)^T
\]

\[
a^{(i-1)} = [a_1, a_2, \ldots, a_{i-1}]
\]

All the points \(x\) can obtain their low dimensional representations \(Y\) through the linear orthogonal transformation vector \(a\), that is \(Y = a^T x\). The ONPE method inherits the power of preserving the local neighborhood relationship and the nonlinear dimension reduction from LLE. Besides, the orthogonal constraint can avoid the singularity problems in data reconstruction.

3. KERNEL ORTHOGONAL NEIGHBORHOOD PRESERVING EMBEDDING

Generally, the nonlinear data in an input space is more likely linear separable after high dimensional nonlinear mapping. In this section, we extend ONPE by the kernel trick, named kernel orthogonal neighbourhoood preserving embedding (KONPE). In KONPE, the original data are mapped into the high dimensional feature space, where the new dataset may present linear relation and make the linear ONPE computationally tractable.

Given a set of cantered observation \(X(x_1,\ldots,x_n) \in \mathbb{R}^n\). By a nonlinear mapping \(\Phi\), the input data are mapped into the high dimensional feature space \(\Phi(x_1),\ldots,\Phi(x_n)\). Assume that the data have been centred so that they satisfy \(\sum_i \Phi(x_i) = 0\). Let \(v(v_1,\ldots,v_d)\) be the linear transformation vector to project the data from the feature space \(\Phi(x)\) into the project space, that is: \(y = v^T \Phi(x)\). Then the objective function of KONPE becomes:
We can notice that, to solve above problems in (12), the inverse of the matrix $K K$ should compute. $K K$ can be singular as the kernel matrix as $K$ is often singular (Scholköpf et al., 2001). The singular regularization is used to ensure the nonsingularity of $K K$, that is:

$$\mathbf{K K} := \mathbf{K K} + \delta \mathbf{I}_n$$

(18)

Where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\delta > 0$ is called the regularization parameter. The KONPE can capture the nonlinear structure in data that is missed by the linear ONPE.

### 4. Fault Detection Based on ONPE and KONPE

In this section, fault detection methods are developed based on ONPE and its kernel generalization, respectively. Referring to the traditional PCA based method; we first construct the ONPE and KONPE model based on the normal operating samples. The square prediction error (SPE) and Hotelling $T^2$ statistics and their confidence limits are then constructed. For online monitoring, the two statistics are used to detect whether the subsequent process is in control or not.

Given the input sample $\mathbf{X}(x_1, \ldots, x_n)$, it is projected into the low dimensional vector $\mathbf{Y}(y_1, \ldots, y_d)$ ($d \leq m$). In the ONPE model the relationship between $\mathbf{X}$ and $\mathbf{Y}$ is $\mathbf{Y} = \mathbf{A X}$ by the linear transformation matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$, While in the KONPE the $\mathbf{Y}$ is obtained by $\mathbf{Y} = \mathbf{a K} (\mathbf{X, X})$. When a new sample $\mathbf{x}_{\text{new}}$ has been collected from the process, the projection $\mathbf{y}_{\text{new}}$ is calculated from the ONPE model by $\mathbf{y}_{\text{new}} = \mathbf{a X}_{\text{new}}$, while in the KONPE it is computed by $\mathbf{y}_{\text{new}} = \mathbf{a K} (\mathbf{x}_{\text{new}})$.

The variation with the model is measured by the $T^2$ statistic, which is defined by:

$$T^2 = \mathbf{y}_{\text{new}}, \mathbf{A}^{-1} \mathbf{y}_{\text{new}}$$

(19)

Where and $\mathbf{A}^{-1}$ is the inverse of the covariance matrix of the $\mathbf{Y}$ of the normal operating conditions in the model space, that is $\mathbf{A}^{-1} = (\mathbf{Y}^\top \mathbf{Y} / (n-1))^{-1}$.

The confidence limit for $T^2$ can be approximated by means of an $F$-distribution with level of significance $\alpha$:

$$T^2_{(1-\alpha)} \sim \frac{d(n-1)}{n-d} F(d, n-d; \alpha)$$

(20)

The SPE statistic is used for monitoring the noisy part of the process information, and also the measurement of how well the sample conforms to the model. To obtain the SPE statistic, we should reconstruct the vector in the input space. The $\mathbf{x}_{\text{new}}$ is reconstructed in the ONPE model by $\mathbf{x}_{\text{new}} = \mathbf{y}_{\text{new}} \mathbf{A}^\top$. The SPE statistic of the ONPE model can be calculated by the following:

$$\text{SPE} = (\mathbf{x}_{\text{new}} - \mathbf{\hat{x}}_{\text{new}}) (\mathbf{x}_{\text{new}} - \mathbf{\hat{x}}_{\text{new}})^\top$$

(21)

While in KONPE, with the reconstructed feature vector $\Phi \mathbf{x}(x)$ in the feature space we have the SPE computed as follows:

$$\text{SPE} = \|\Phi \mathbf{x}(x) - \Phi \mathbf{\hat{x}}(x)\|^2 = \mathbf{K}(x, x) - \sum_{j=1}^{m} y_{\text{new}}^2$$

(22)

The confidence limit for the SPE statistic can be determined by a $\chi^2$ distributed function.
\[ SPE \sim g \cdot \chi^2_k \] (23)

Where \( g \) and \( h \) are the parameter of \( \chi^2 \) distribution, which can be approximated by \( g = \frac{b}{2a} \) and \( h = \frac{a^2}{2m} \) respectively. \( a \) and \( b \) are the estimated mean and variance of the SPE from the reference process.

It points out that the dot products are represented by the kernel function, which plays a central role in the KONPE. There are a number of kernel functions; in the paper we select the representative radial basis function (RBF) which is said the best for the nonlinear process monitoring:

\[
k(x,y) = \exp(-\frac{\|x-y\|^2}{\sigma})
\]

(24)

The kernel parameter \( \sigma \) is specified a priori by the user that determined empirically.

5. SIMULATION RESULTS AND DISCUSSION

In order to analyze the fault detection performance of the two proposed methods, two system simulation studies are conducted. One is the Tennessee Eastman (TE) benchmark process, and the other is a three-variable nonlinear system introduced by Dong and McAvoy. These two simulations have been widely accepted for testing various monitoring approaches. To test the superiority of the methods, the KPCA model is also constructed and comparative studies are carried out among different methods. The \( T^2 \) and SPE monitoring statistics are constructed, with the confidence limits is set as 99%. The number of nearest neighbours for KONPE and ONPE is set as 12, and the regularization parameter is set as \( \delta = 10^{-6} \).

5.1 Tennessee Eastman process

Tennessee Eastman (TE) process is proposed by Downs and Vogel based on a realistic standard chemical industrial process model (Downs et al., 1993). The model is including five major unit operations: a reactor, a condenser, a flash separator, a stripper and a recycle compressor. The process with 41 measured variables and 12 manipulated variables can simulate a wide variety of normal operating conditions, together with 21 programmed faulty conditions. The data generated in the process with the characteristics of nonlinearity, strong coupling and dynamical. As a standard benchmark simulation for the complex industrial process, the simulator has been extensively used for comparing various process monitoring strategies and control schemes (Ge et al., 2009; Shao et al., 2009).

Both the training data and test data are acquired from the TE process, consisting of 960 samples, and their sampling interval is 3 minutes. The training data are collected under normal operating conditions and the test data sets are obtained with the fault introduced at sample 161. The dimension of the model space for the three methods is chosen as 20, which explains 97.41% of the total process information. KPCA and KONPE reaches their best performances at kernel parameter \( \sigma = 35 \). Three monitoring models of KPCA, KONPE and ONPE are developed based on the training dataset. Monitoring results of some representative faults are shown in Fig.1, Fig.2 and Fig.3. Twenty Monte Carlo experiment results of the missing alarm rate for all 21 fault modes are shown in Table 1. The smaller value in the table, the better performance of the corresponding statistic shows, and the best achieved performance for each fault is marked with a bold number.

Table 1: the false alarm rate and missing alarm rate on TE benchmark

<table>
<thead>
<tr>
<th>Fault</th>
<th>KPCA T² SPE</th>
<th>KONPE T² SPE</th>
<th>ONPE T² SPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5 0.2</td>
<td>0.2 0.2</td>
<td>0.5 0.13</td>
</tr>
<tr>
<td>2</td>
<td>2.12 1.5</td>
<td>1.5 1.75</td>
<td>1.88 3.13</td>
</tr>
<tr>
<td>3</td>
<td>99.8 94.5</td>
<td>84.8 98.1</td>
<td>98.4 98.1</td>
</tr>
<tr>
<td>4</td>
<td>75.8 0</td>
<td>4 25.5</td>
<td>66.1 0</td>
</tr>
<tr>
<td>5</td>
<td>82 68.3</td>
<td>33 74</td>
<td>0 84</td>
</tr>
<tr>
<td>6</td>
<td>1 0</td>
<td>0 0</td>
<td>0 1</td>
</tr>
<tr>
<td>7</td>
<td>0 0</td>
<td>0 0</td>
<td>1 0</td>
</tr>
<tr>
<td>8</td>
<td>10 2</td>
<td>1 2</td>
<td>2 9</td>
</tr>
<tr>
<td>9</td>
<td>100 96</td>
<td>88 98</td>
<td>98 99</td>
</tr>
<tr>
<td>10</td>
<td>85 43</td>
<td>36 57</td>
<td>12 92</td>
</tr>
<tr>
<td>11</td>
<td>59 24</td>
<td>29 42</td>
<td>55 26</td>
</tr>
<tr>
<td>12</td>
<td>8 1</td>
<td>0 1</td>
<td>0 10</td>
</tr>
<tr>
<td>13</td>
<td>3 0</td>
<td>0 0</td>
<td>0 1</td>
</tr>
<tr>
<td>14</td>
<td>100 91</td>
<td>83 96</td>
<td>96 98</td>
</tr>
<tr>
<td>15</td>
<td>94 52</td>
<td>41 75</td>
<td>8 96</td>
</tr>
<tr>
<td>16</td>
<td>27 4</td>
<td>6 13</td>
<td>18 6</td>
</tr>
<tr>
<td>17</td>
<td>12 10</td>
<td>9 11</td>
<td>11 10</td>
</tr>
<tr>
<td>18</td>
<td>96.6 85.5</td>
<td>73 96.7</td>
<td>9.5 71.75</td>
</tr>
<tr>
<td>19</td>
<td>86.6 43.8</td>
<td>31.25 55</td>
<td>8.88 56.13</td>
</tr>
<tr>
<td>20</td>
<td>58 55</td>
<td>50 60</td>
<td>55 54</td>
</tr>
<tr>
<td>21</td>
<td>58 55</td>
<td>50 60</td>
<td>55 54</td>
</tr>
</tbody>
</table>

From the table, it’s evident that the ONPE and KONPE take up the absolute superiority. The three methods provide similar performances for faults that are easy to detect. All the monitoring schemes provide high missing alarm rates for the faults that are relatively difficult to detect, the faults 3, 5, 9, 10, 15, 16, 19, 20. However, the ONPE and KONPE provide notably lower missing alarm rate than KPCA for these faults. Note that the KONPE exhibits the significant improvements in the \( T^2 \) statistic for most of the faults, which can detect the fault accurate and timely with minimal missing alarm rate. It can be inferred that, by preserving the local manifold among the data, the KONPE can extract more useful information from the process dataset and has a higher sensitivity to detect these faults than KPCA. The linear methods are more appropriate for some faults, like faults 19 and 20; as ONPE has little better performance than KONPE.

Fig.1. The \( T^2 \) and SPE monitoring charts with 99% confidence limits for fault 4 in TE process.
5.2 Three-variables nonlinear system

Consider the following nonlinear numerical example composed of three variables with one degree of freedom:

\[ x_1 = t + e_1 \]
\[ x_2 = t^2 - 3t + e_2 \]
\[ x_3 = -t^3 + 3t^2 + e_3 \] \hspace{1cm} (25)

Where \( e_1, e_2, \) and \( e_3 \) are independent noise of \( N(0, 0.01) \) and \( t \in [0.01, 2] \). 500 normal samples were generated by equations (25) as the training samples. Three additional test data sets with 500 samples were also generated according to the following artificial fault respectively:

(a) A step change by -1 was added to \( x_2 \) from sample 251.
(b) A linear increase by 0.005\((k-5)\) was added to \( x_1 \) from sample 251, where \( k \) is the sample number.
(c) A step change by -0.5 was added to \( t \) from sample 251.

After the training samples have been collected and scaled to zero mean and unit variance, the traditional KPCA, the new proposed KONPE and the linear ONPE model are developed. The KPCA and KONPE reach their best performances at kernel parameter \( \sigma \) determined as 20, which is suggested 5\( n \) ~ 10\( n \) (\( n \) is the dimension of the input space) (Lee et al., 2004). The dimension of reduced space is chosen as 2, which can explain 95.34% of total variation in the training data. The monitoring results of the three methods for the fault process are illustrated in Fig.4–Fig.6, are also with the 99% confidence limits. 20 times Monte Carlo experiment of the missing alarm rate and false alarm rate of each statistic are listed are listed in Table 2. The smaller value in the table, the better performance of the corresponding statistic shows.

The monitoring results show that the KONPE outperforms other methods with relatively lower missing alarm rate and false alarm rate. Especially the \( T^2 \) statistics is much more sensitive to the faults than the corresponding statistics of KPCA and ONPE. KPCA has the worst monitoring performance in the \( T^2 \) statistic, which cannot detect the fault \( a \) and fault \( b \) effectively, and shows high missing alarm rate for the fault \( c \). This is because the KONPE and ONPE have greater superiority in preserving the local neighbour structure information between points than KPCA, thus it succeeds in identifying the underlying geometrical structure and better embodying the process change in the feature space. The KONPE exhibits superiority in contrast to the linear ONPE monitoring, which verified the efficiency of the KONPE in capture the nonlinear relationship in the process variables.
Table 2: The missing alarm rate and false alarm rate for the simple system

<table>
<thead>
<tr>
<th>Fault</th>
<th>Missing alarm rate (%)</th>
<th>False alarm rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>KPCA</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>SPE</td>
<td>1.44</td>
<td>31.9</td>
</tr>
<tr>
<td>KONPE</td>
<td>27.8</td>
<td>74</td>
</tr>
<tr>
<td>SPE</td>
<td>2.94</td>
<td>61.2</td>
</tr>
<tr>
<td>ONPE</td>
<td>37.2</td>
<td>87.1</td>
</tr>
<tr>
<td>SPE</td>
<td>74.2</td>
<td>48.4</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

In this paper, the ONPE method has been applied for nonlinear process monitoring and extended to its kernel form. Compared to some global modelling approaches such as PCA and KPCA, the local preserving feature of ONPE enables it to find more meaningful data features that may be hidden in the high dimensional observations. Inherently, the ONPE method has the ability of nonlinear dimension reduction in LLE and the additional orthogonal characteristic. Therefore, the data reconstruction becomes available in this method, which is very useful to carry out online process monitoring. Both of the ONPE and KONPE models have been used for nonlinear modelling and fault detection purposes. Simulation results of the TE benchmark process and a nonlinear system have illustrated the superiority of the ONPE-based methods in comparison with the widely used KPCA based method.

ACKNOWLEDGEMENT

This work was financially supported by the National Natural Science Foundation of China (No. 60974056) and National Basic Research Program of China (973 Program, Grant No. 2012CB720505).

REFERENCES


