Journal of Process Control 61 (2018) 1–11

Contents lists available at ScienceDirect

Journal of Process Control

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

Online reduced kernel principal component analysis for process monitoring

Radhia Fezai[a], Majdi Mansouri[b], Okba Taouali[a,*], Mohamed Faouzi Harkat[c], Nasreddine Bouguila[a]

[a] Laboratory of Automatic Signal and Image Processing, National School of Engineers of Monastir, University of Monastir, 5019, Tunisia
[b] Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar
[c] Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar

ARTICLE INFO

Article history:
Received 17 October 2016
Received in revised form
27 September 2017
Accepted 18 October 2017
Available online 21 November 2017

Keywords:
Principal component analysis (PCA)
Kernel PCA
Reduced kernel PCA
Dictionary
Dynamic process
Fault detection

ABSTRACT

Kernel principal component analysis (KPCA), which is a nonlinear extension of principal component analysis (PCA), has gained significant attention as a monitoring method for nonlinear processes. However, KPCA cannot perform well for dynamic systems and when the training data set is large. Therefore, in this paper, an online reduced KPCA algorithm for process monitoring is proposed. The process monitoring performances are studied using two examples: a numerical example and Tennessee Eastman Process (TEP). The simulation results demonstrate the effectiveness of the proposed method when compared to the online KPCA method.

1. Introduction

The last century has seen an increasingly emergence of manufacturing and process industries, with the rise of energy costs and environmental regulations. In these environments, processes are highly automated. Therefore, the monitoring algorithms are important to detect any fault that might occur. Multivariate statistical methods such as principal component analysis (PCA) [8,16,3,27], independent component analysis (ICA) [17,20,10] and partial least squares (PLS) [28,1] have been applied for process monitoring. PCA is the most used technique because of the simplicity of its concept and implementation. It transforms the original input variables into variables in a low dimensional subspace. Process monitoring is detecting faults in a system. In the PCA method, fault detection is performed using fault detection indices. The most used indices are the squared prediction error (*SPE*) that monitors the residual subspace and the Hotelling T^2 statistic for monitoring the principal component subspace. The PCA method assumes that the relationships among variables are linear and may ignore important information in the behaviors of nonlinear systems. However, the majority of industrial processes have nonlinear characteristics. To address this issue, many nonlinear PCA methods have been developed [18,6,26]. The most popular one is kernel principal component analysis (KPCA) [26,23] because of its simplicity and elegance. It has been shown in [21], that KPCA method is used for monitoring continuous processes and it provided better performance when compared to linear PCA method. KPCA can efficiently compute principal components in high-dimensional feature space using integral operators and nonlinear kernel functions. It consists in mapping measurements from their original space to a higher dimensional feature space where PCA is performed. Fault detection using KPCA [22,4,24,9] can be performed similarly to linear PCA using *SPE* and T^2 statistics in the feature space. In the KPCA method, the size of the kernel matrix is given by the number of the training samples. Every time a new sample x is collected, N kernel functions have to be evaluated to form the kernel vector $k(x)$. Therefore, using KPCA model for process monitoring imposes a high computational cost when the training data set is large. K-means clustering [25] was developed in order to reduce the amount of training data. It requires the number of clusters in advance, therefore using it for process monitoring leads to numerous errors because it doesn't consider the variation of parameters according to the process operation changes.

In [15], the fault detection singular value decomposition reduced kernel principal component analysis (SVD-RKPCA) method

* Corresponding author.
E-mail address: taoualiok@gmail.com (O. Taouali).

https://doi.org/10.1016/j.jprocont.2017.10.010
0959-1524/

is proposed for online monitoring of nonlinear processes. The SVD-RKPCA method is performed in two phases (offline and online). It consists in finding a reduced data set by selecting the variables that have the highest projection variance to build the initial KPCA model in the offline phase and then updating this model using the SVD-KPCA technique in the online phase. In the SVD-RKPCA method, the dictionary size is fixed during the training and testing phases. Thus, the large size of reduced data set can lead to a significant reduction in the computational efficiency. In addition, when the size of reduced data set is small, the updated information on the systems can be ignored and SVD-RKPCA cannot perform well for dynamic systems. To overcome the problems stated above, a new method referred as online reduced KPCA for process monitoring is proposed. The method aims to reduce the number of kernel functions. The reduced set of kernel functions is called dictionary. A new criterion based on squared prediction error evaluation determines the relevance of discarding or adding the kernel function to the dictionary $\mathcal{D}_r$. The resulting dictionary contains linearly independent kernel functions. The proposed method is performed in the online phase and takes into consideration the dynamic behaviors of the systems by changing the model structure. In the proposed online reduced KPCA method, the size of dictionary is variable and depends on the evaluated criterion. The proposed method forms the dictionary in an online way and then it updates the old KPCA model and uses it for process monitoring. The new method reduces significantly the computation time required to detect faults in nonlinear systems while conserving the structure of the data in the feature space. The performance of the proposed method is evaluated using a numerical example and Tennessee Eastman Process (TEP). The simulation results show that the proposed method provides better detection performances when compared to the online KPCA method.

This paper is organized as follows: an overview of the PCA and KPCA methods is given in Section 2. The fault detection index *SPE* is presented in Section 3. Section 4 presents the proposed online reduced KPCA method. Then, in Section 5, the fault detection performance is studied using a numerical example and TEP. At the end, the conclusions are presented in Section 6.

2. Preliminaries

2.1. Principal component analysis

The main idea of the principal component analysis (PCA) method is discarding the noise and collinearity between process variables, while preserving the most important information of the original data set. A PCA model is established based on the data collected under normal operating conditions for process monitoring.

Let $X \in \mathbf{R}^{N\times m}$ be the data matrix with N measurements and m variables. For PCA modeling, the matrix X is scaled to zero mean and unit variance. The covariance matrix Σ of the data matrix X and its eigenvalue decomposition are given by:

$$\begin{aligned}\Sigma &= \frac{1}{N}X^TX \\ &= \begin{bmatrix}\hat{P} & \tilde{P}\end{bmatrix}\begin{bmatrix}\hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda}\end{bmatrix}\begin{bmatrix}\hat{P} & \tilde{P}\end{bmatrix}^T\end{aligned} \tag{1}$$

where $\hat{P} \in \mathbf{R}^{m\times\ell}$ and $\tilde{P} \in \mathbf{R}^{m\times(m-\ell)}$, representing the first ℓ and last $(m-\ell)$ eigenvectors of Σ, respectively. ℓ is the number of retained principal components (PCs). The diagonal matrices $\hat{\Lambda}$ and $\tilde{\Lambda}$ contain the ℓ highest and $(m-\ell)$ lowest eigenvalues of Σ, respectively. The data matrix X can be decomposed as follows:

$$X = \hat{X} + \tilde{X} \tag{2}$$

where $\hat{X}$ and $\tilde{X}$ are the projection to the principal subspace (PS), spanned by the columns of $\hat{P}$, and to the residual subspace (RS) spanned by the columns of $\tilde{P}$.

2.2. Kernel principal component analysis

Since the classical PCA method performs well only on linear processes, a nonlinear PCA technique called kernel PCA (KPCA), has been developed by Schölkopf et al. [26] and widely used to model various nonlinear processes. The key idea of KPCA is to map first the input space into a feature space $\mathcal{H}$ via nonlinear mapping ϕ and then perform a linear PCA in $\mathcal{H}$. Let the normalized training set be $x_1, x_2, \ldots, x_N \in \mathbf{R}^m$, The measured inputs are projected into the feature space using the mapping function ϕ:

$$\phi : x_i \in \quad \mathbf{R}^m \mapsto \phi(x_i) \in \mathbf{R}^h \tag{3}$$

where $h \gg N$ is the dimension of the feature space. An important property of the feature space is that the dot product of two vectors $\phi(x_i)$ and $\phi(x_j)$, $i, j = 1, \ldots, N$, can be calculated as:

$$\phi(x_i)^T\phi(x_j) = \langle\phi(x_i), \phi(x_j)\rangle = k(x_i, x_j) \tag{4}$$

where k is called the kernel function. There are several types of kernel functions. One of the most used kernel functions is the radial basis function (RBF) which is defined as:

$$k(x_i, x_j) = \exp\left[\frac{-\|x_i - x_j\|^2}{\sigma^2}\right] \tag{5}$$

where σ is the width of a Gaussian function. Assuming that the vectors in the feature space are scaled to zero mean, the mapped data in $\mathcal{H}$ is arranged as

$$\mathcal{X} = \begin{bmatrix}\phi(x_1) & \phi(x_2) & \cdots & \phi(x_N)\end{bmatrix}^T.$$

The covariance matrix Q in the feature space $\mathcal{H}$ can be calculated as:

$$\begin{aligned}Q &= \frac{1}{N-1}\mathcal{X}^T\mathcal{X} \\ &= \frac{1}{N-1}\sum_{j=1}^{N}\phi(x_j)\phi(x_j)^T\end{aligned} \tag{6}$$

where it is assumed that $\sum_{j=1}^{N}\phi(x_j) = 0$. Similarly to the linear PCA, KPCA in the feature space is equivalent to solving the following eigenvalue problem:

$$\begin{aligned}\mu_k V_k &= QV_k \\ &= \frac{1}{N-1}\sum_{j=1}^{N}\langle\phi(x_j), V_k\rangle\phi(x_j)\end{aligned} \tag{7}$$

where μ_k and V_k represent respectively the kth eigenvalue and eigenvector of Q, and $\langle .,. \rangle$ is the dot product. For $\mu_k \neq 0$, it is clear from Eq. (7) that every eigenvector V_k of Q can be considered as a linear combination of $\phi(x_1), \ldots, \phi(x_N)$. Thus, there are coefficients $\alpha_{k,i}\,(i = 1, \ldots, N)$ such that:

$$\begin{aligned}V_k &= \sum_{i=1}^{N}\alpha_{k,i}\phi(x_i) \\ &= \mathcal{X}^T\alpha_k\end{aligned} \tag{8}$$