Control Engineering Practice 80 (2018) 17–25

Contents lists available at ScienceDirect

Control Engineering Practice

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

Parallel PCA–KPCA for nonlinear process monitoring

Qingchao Jiang, Xuefeng Yan *

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

ARTICLE INFO

Keywords:
Nonlinear process monitoring
Fault detection
Parallel PCA–KPCA
Randomized algorithm
Genetic algorithm

ABSTRACT

Both linear and nonlinear relationships may exist among process variables, and monitoring a process with such complex relationships among variables is imperative. However, individual principal component analysis (PCA) or kernel PCA (KPCA) may not be able to characterize these complex relationships well. This paper proposes a parallel PCA–KPCA (P-PCA–KPCA) modeling and monitoring scheme that incorporates randomized algorithm (RA) and genetic algorithm (GA) for efficient fault detection for a process with linearly correlated and nonlinearly related variables First, to determine the included variables in the parallel PCA (P-PCA) and the parallel KPCA (P-KPCA) models, GA-based optimization is performed, in which RA is used to generate faulty validation data. Second, monitoring statistics are established for the P-PCA and the P-KPCA models, in which the process status is determined. The proposed monitoring scheme discriminates the linear and nonlinear relationships among variables in a process and deals with nonlinear processes efficiently. We provide case studies on a numerical example and the continuous stirred tank reactor process. These case studies demonstrate that the proposed P-PCA–KPCA monitoring scheme is better than conventional PCA- or KPCA-based methods at performing nonlinear process monitoring.

1. Introduction

Process monitoring that focuses on fault detection and diagnosis plays an important role in ensuring process safety (Chiang, Russell, & Braatz, 2001; Jiang & Huang, 2016; Zhao & Gao, 2017). In recent years, data-based fault detection and diagnosis methods have become a hot topic because of the rapid advancement of data gathering and storing techniques (Chen, 2016; Ding, 2014; Jiang, Ding, Wang, & Yan, 2017; Ma, Dong, Peng, & Zhang, 2017; Rashidi, Singh, & Zhao, 2017; Wang & Zhao, 2017). As one of the most fundamental techniques, principal component analysis (PCA) has been intensively studied and extended (Bakdi, Kouadri, & Bensmail, 2017; Jiang & Yan, 2014). Dynamic PCA was developed to deal with dynamic processes with time-series autocorrelation (Ku, Storer, & Georgakis, 1995; Li, Qin, & Zhou, 2014). Recursive or moving window PCA was developed for time-varying processes (Li, Yue, Valle-Cervantes, & Qin, 2000; Wang, Kruger, & Irwin, 2005). Multiway PCA was applied in monitoring batch processes (Nomikos & Macgregor, 1994; Wise, Gallagher, Butler, White, & Barna, 2015). Distributed PCA was developed for processes that consist of a large number of variables (Ge & Song, 2013; Jiang, Yan, & Huang, 2016). Although numerous successful applications have been reported, PCA performs poorly in dealing with nonlinear processes because it characterizes only the linear correlation among variables and does not explore the nonlinear relationships.

Several nonlinear monitoring methods have been proposed to deal with the nonlinearity of a process. Neural network-based approaches were developed (Kramer, 1991; Yuan et al., 2017). However, nonlinear optimization is generally involved in these methods (Dong & McAvoy, 1996; Kramer, 1991). To avoid nonlinear optimization, kernel PCA (KPCA) was proposed and extended intensively (Lee, Yoo, Sang, Vanrolleghem, & Lee, 2004; Schölkopf, Smola, & Müller, 1998). For example, an adaptive KPCA monitoring method was proposed in Cheng, Hsu, and Chen (2010), a KPCA combined with kernel density estimation (KDE) method was proposed in Samuel and Cao (2016), and a multivariate statistical KPCA method was presented in Luo, Li, Deng, Zhong, and Cai (2016). Despite successful applications of KPCA (Jiang & Yan, 2015; Yi et al., 2017; Zhang, Du, & Li, 2017), some issues remain unresolved, such as the monitoring of a process with both linearly correlated and nonlinearly related variables. Given the large number of variables, including all variables in one KPCA model is not appropriate because doing so prevents the data structure from being characterized well. The monitoring of a chemical process with linear and nonlinear relationships remains an open question.

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

https://doi.org/10.1016/j.conengprac.2018.07.012
Received 2 March 2018; Received in revised form 3 June 2018; Accepted 27 July 2018

Hybrid modeling techniques have attracted considerable attention in the area of process modeling and monitoring (Chen, 2011; Zhang, Chai, & Wang, 2017). The key idea of hybrid modeling is to use different types of models or approaches to characterize different relationships among variables. Several results for hybrid nonlinear process modeling have been reported (Chen, 2011; Zhang, Chai, & Wang, 2017). Recently, a serial PCA (SPCA)-based modeling and monitoring method that performs KPCA using residual PCA has been proposed (Deng, Tian, Chen, & Harris, 2016). Although the efficiency of this method has been demonstrated, the nonlinear relation among measured variables may be concealed in the extracted features of PCA, and the linear and nonlinear relationships may not be well represented. Moreover, the monitoring performance may be degraded.

Inspired by the idea of hybrid modeling, this paper proposes a parallel PCA–KPCA (P-PCA–KPCA) modeling and monitoring scheme to deal with processes that contain linearly correlated and nonlinearly related variables. The key issue is to explore the complex relationships among variables and determine which variable should be included in the PCA or the KPCA models, i.e., variable emplacement. This paper proposes an randomized algorithm (RA) integrated with a genetic algorithm (GA)-based method to automatically emplace process variables. RA has been widely used to deal with uncertain issues in robust control (Karp, 1991; Tempo, Calafiore, & Dabbene, 2005) as well as fault diagnosis scheme design (Chen, Ding, Peng, Yang, & Gui, 2017; Jiang, Wang, & Yan, 2017; Krueger et al., 2017). Here RA is employed to generate faulty validation data, based on which the GA optimization is performed. To the best of our knowledge, this is the first time that a P-PCA–KPCA fault detection model is established. Unlike classical PCA and KPCA methods, the proposed P-PCA–KPCA monitoring scheme discriminates linear and nonlinear relationships and therefore exhibits superiority in dealing with nonlinear processes.

The remainder of this paper is structured as follows: Section 2 provides a brief review of the basics of PCA, KPCA, and RA. The monitoring problem of nonlinear processes with linear and nonlinear relationships is formulated. Section 3 details the proposed P-PCA–KPCA monitoring scheme. Section 4 presents application examples on a numerical nonlinear example and continuously stirred tank reactor (CSTR) process and shows performance comparisons with some state-of-the-art methods. Section 5 draws the conclusions.

2. Preliminaries and problem formulation

2.1. PCA-based fault detection

PCA is generally used to deal with a linear process with Gaussian-distributed variables. Let $\boldsymbol{X} \in \mathbb{R}^{m\times N}$ denote a set of normalized data with m variables and N samples. Performing singular value decomposition on the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$ derives (Ding, 2014)

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T, \tag{1}$$

where $\boldsymbol{\Lambda} = diag\left(\lambda_1, \ldots, \lambda_m\right)$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$. In PCA, the load matrix $\boldsymbol{P} \in \mathbb{R}^{m\times m}$ is divided as $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{P}_{pc} & \boldsymbol{P}_{res}\end{bmatrix}$ and $\boldsymbol{\Lambda}$ is divided as $\boldsymbol{\Lambda} = \begin{bmatrix}\boldsymbol{\Lambda}_{pc} & 0\\ 0 & \boldsymbol{\Lambda}_{res}\end{bmatrix}$. Given a data sample $\boldsymbol{x} \in \mathbb{R}^{m\times 1}$, two statistics, namely, T^2 and Q, are constructed as (Ding, 2014)

$$T^2 = \boldsymbol{x}^T\boldsymbol{P}_{pc}\boldsymbol{\Lambda}_{pc}^{-1}\boldsymbol{P}_{pc}^T\boldsymbol{x}, \tag{2}$$

$$Q = \left\|\left(\boldsymbol{I} - \boldsymbol{P}_{pc}\boldsymbol{P}_{pc}^T\right)\boldsymbol{x}\right\|_E^2 = \boldsymbol{x}^T\left(\boldsymbol{I} - \boldsymbol{P}_{pc}\boldsymbol{P}_{pc}^T\right)\boldsymbol{x}, \tag{3}$$

where $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ is the identity matrix.

2.2. KPCA-based fault detection

KPCA is generally used to deal with nonlinear processes. Let $\boldsymbol{x}_j \in \mathbb{R}^m, j = 1, \ldots, N$, denote a set of zero-mean data. Through nonlinear mapping $\boldsymbol{\Phi}(\cdot)$, the covariance of the mapped data in the feature space $\boldsymbol{C}^F$ is (Lee et al., 2004; Schölkopf et al., 1998)

$$\boldsymbol{C}^F = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)^T. \tag{4}$$

$\boldsymbol{C}^F$ can be diagonalized by eigenvalue decomposition as

$$\lambda\boldsymbol{v} = \boldsymbol{C}^F\boldsymbol{v}, \tag{5}$$

where $\lambda \geq 0$ denotes the eigenvalues, and $\boldsymbol{v}$ denotes the eigenvectors. Substituting Eq. (4) in Eq. (5) derives (Lee et al., 2004; Schölkopf et al., 1998)

$$\boldsymbol{C}^F\boldsymbol{v} = \left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)^T\right)\boldsymbol{v} = \frac{1}{N}\sum_{j=1}^{N}\left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{v}\right\rangle\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right). \tag{6}$$

Considering that all solutions $\boldsymbol{v}$ with $\lambda \neq 0$ lie in the span of $\boldsymbol{\Phi}\left(\boldsymbol{x}_1\right), \ldots, \boldsymbol{\Phi}\left(\boldsymbol{x}_N\right)$, coefficients $\alpha_i (i = 1, \ldots, N)$ that satisfy $\boldsymbol{v} = \sum_{i=1}^{N}\alpha_i\boldsymbol{\Phi}\left(\boldsymbol{x}_i\right)$ exist. Then, Eq. (5) becomes (Lee et al., 2004; Schölkopf et al., 1998)

$$\lambda\sum_{i=1}^{N}\alpha_i\left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_k\right), \boldsymbol{\Phi}\left(\boldsymbol{x}_i\right)\right\rangle = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_k\right), \sum_{j=1}^{N}\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\right\rangle\left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{\Phi}\left(\boldsymbol{x}_i\right)\right\rangle \tag{7}$$

for all $k = 1, \ldots, N$. With the introduction of a kernel matrix $\boldsymbol{K}$ with $[\boldsymbol{K}]_{ij} = K_{ij} = \left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_i\right), \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\right\rangle$, Eq. (7) becomes (Lee et al., 2004; Schölkopf et al., 1998)

$$\lambda N\boldsymbol{K}\boldsymbol{\alpha} = \boldsymbol{K}^2\boldsymbol{\alpha}, \tag{8}$$

where $\boldsymbol{\alpha} = \left[\alpha_1, \ldots, \alpha_N\right]^T$. For a new sample $\boldsymbol{x}$, the kernel principal component (KPC) is calculated as (Lee et al., 2004; Schölkopf et al., 1998)

$$t_k = \left\langle\boldsymbol{v}_k, \boldsymbol{\Phi}(\boldsymbol{x})\right\rangle = \sum_{i=1}^{N}\alpha_i^k\left\langle\boldsymbol{\Phi}\left(\boldsymbol{x}_i\right), \boldsymbol{\Phi}(\boldsymbol{x})\right\rangle, \tag{9}$$

where $k = 1, \ldots, p$ (p denotes the number of retained KPCs). The statistics of T^2 and Q are formulated as (Lee et al., 2004; Schölkopf et al., 1998)

$$T^2 = [t_1, \ldots, t_p]\boldsymbol{\Lambda}^{-1}[t_1, \ldots, t_p]^T, \tag{10}$$

$$Q = \left\|\boldsymbol{\Phi}(x) - \hat{\boldsymbol{\Phi}}_p(x)\right\|^2 = \sum_{j=1}^{n}t_j^2 - \sum_{j=1}^{p}t_j^2, \tag{11}$$

where n denotes the number of non-zero eigenvalues.

2.3. Randomized algorithm

In the design of a process monitoring scheme, false alarm rate (FAR) and fault detection rate (FDR) are two important indices for evaluating the monitoring performance. The computation of FAR and FDR involves the computation of probability where $J > J_{th}$ under certain (fault) conditions. Given the accuracy requirement $\varepsilon \in (0, 1)$ and the confidence level $\delta \in (0, 1)$, RA delivers an estimate $\hat{p}(\gamma)$ of $p(\gamma) = prob(J(\omega) \leq \gamma)$ (the probability of $J(\omega) \leq \gamma$) such that

$$p(\gamma) < \hat{p}(\gamma) + \varepsilon \tag{12}$$

with a probability of at least $1 - \delta$, where ω is the random variable with the known density $D(\omega)$ and support D_ω (Tempo et al., 2005). In the RA framework, N_R independent identical distributed random