Chemometrics and Intelligent Laboratory Systems 127 (2013) 121–131

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical 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

Article history:
Received 7 March 2013
Received in revised form 19 June 2013
Accepted 19 June 2013
Available online 1 July 2013

Keywords:
Weighted kernel principal component analysis
Nonlinear process monitoring
Probability density estimation
Moving window

ABSTRACT

Kernel principal component analysis (KPCA) has been widely used in nonlinear process monitoring; however, KPCA does not always perform efficiently because useful information may be submerged under retained KPCs. To address this shortcoming, probability density estimation- and moving weighted window-based KPCA (PM-WKPCA) is proposed. PM-WKPCA is used mainly to estimate the probability and evaluate the importance of each KPC by kernel density estimation and then set different weighting values on KPCs to highlight the useful information. The status of the process is also evaluated comprehensively using weighted statistics within a moving window. The efficiency of the proposed method is demonstrated by the following: case studies on a numerical nonlinear system, the simulated continuously stirred tank reactor process, and the Tennessee Eastman process. Monitoring results indicate that the proposed method is superior to the conventional PCA, KPCA, and some typical extension methods.

1. Introduction

Process monitoring has gained increasing interest because of the rising demand in plant safety and product quality. With recent developments in data collection and computing technology, multivariate statistical process monitoring (MSPM) methods [1–14] have progressed quickly [15–17]. And among these MSPM methods, principal component analysis (PCA) usually serves as the most fundamental one and has been studied intensively and used extensively. PCA can be used effectively with high-dimensional, highly correlated data by projecting the data onto a lower dimensional subspace containing sufficient variance information of normal training data. However, PCA methods assume that the relationship between process variables is linear, which can easily change in practice [18–20].

With the nonlinear behavior of the chemical process considered, numerous nonlinear PCA (NLPCA) approaches have been developed [13,20,21], and the kernel PCA (KPCA) [22] method, which can efficiently compute PCs in a high-dimensional feature space using the kernel function, is the most widely used [22]. The key idea of KPCA is to map the data initially into a feature space by nonlinear mapping and then extract the PCs in the high-dimensional feature space. A major advantage of KPCA is that it only requires solving an eigenvalue problem and requires no nonlinear optimization [18,19]. KPCA has gained considerable attention because of its simplicity and efficiency, and it has also been extended to dynamic KPCA [23] and multiscale KPCA [24], among others, to solve various monitoring problems [10,25–27].

Despite sufficient research on KPCA and numerous successful applications, the approach does not always perform efficiently. In KPCA monitoring, the KPCA model is generated from the normal training data. The monitored variables are first nonlinearly mapped into a high-dimensional space through the kernel mapping, and then the kernel principal components (KPCs) are generated as linear combinations of the variables in the high-dimensional feature space. Generally, the first several KPCs are employed to construct the dominant subspace [18,19,28], and correspondingly, T^2 statistic is used to monitor the variation in the dominant subspace. In the T^2 statistic, the employed KPCs are scaled with similar importance degrees [15,28,29]. However, for a definite fault, the fault usually causes one or several monitored variables' variation, and all the employed KPCs are not guaranteed to have the same variation degree [30]. Namely, there are some KPCs with larger variation, and others with less or without variation. In all the employed KPCs, the KPCs with larger variation contain more fault information and are beneficial to find fault timely, and the KPCs with less variation or without variation contain less or no fault information and aren't used to find fault. Thus, for monitoring the definite fault, the importance degrees of all the employed KPCs are usually not the same (as illustrated in the Motivational example section below). Some KPCs would reflect more information of the fault while others would reflect less. When all the employed KPCs are used to monitor process with the same importance by the T^2 statistics, the useful information (contained in the KPCs with larger variation) for detecting and diagnosing the fault may be submerged by the KPCs with less variation and without variation, and

* Corresponding author at: East China University of Science and Technology, P.O. Box 293, MeiLong Road No. 130, Shanghai 200237, PR China. Tel.: +86 2164251036.
E-mail address: xfyan@ecust.edu.cn (X. Yan).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.06.013

poor monitoring performance happens. Therefore, the importance degree of each employed KPC should be evaluated online and the weighting strategy is needed to highlight the fault information.

Considering the issue of highlighting the useful variation information, several strategies used in MSPM from different aspects have been reported. Wold [31] proposed a timely updated model that weights recent observations more heavily than earlier ones. He et al. [32] highlighted the fault information by weighting related variables heavily in kernel Fisher discriminant analysis for fault diagnosis. Ferreira et al. [33] suggested a sample-wise weighted PCA for multicampaign process monitoring. These methods mainly focused on weighting raw measured variables without analyzing the importance of different latent variables. Rashid and Yu [10,27] used the multidimensional mutual information to evaluate the statistical dependency between the independent component subspaces of the normal benchmark and monitored data sets, and constructed a dissimilarity index which considered the online fault information. This method highlighted the fault information in the ICA dominant subspace, however, the behavior of each latent variable was not analyzed. Jiang and Yan [30] proposed weighted principal component analysis (WPCA) to highlight the useful information reflected on the dominant subspace. Online WPCA evaluates the importance of each PC by using the change rate of the T^2 statistic along each PC and setting different values on the PCs to highlight the useful information for process monitoring. This method improved both fault detection and identification performance; however, as is well known, the nonlinear relationships between variables can't be well extracted by linear PCA, and therefore the WPCA is limited to linear processes.

In this article, probability density estimation- and moving weighted window-based KPCA (PM-WKPCA) is proposed to solve the problem on submerged useful information and improve performance in nonlinear chemical process monitoring. To find and stress the informative KPCs in time, the probability density estimation (PDE) method is employed. First, the probability density function of each KPC in normal condition is estimated through kernel density estimation (KDE) with large amount of normal process data. Second, when online monitoring, the probability of each KPC score in the current sample is determined by the previously obtained distribution, and the importance of the corresponding KPC is evaluated. Then, different weighting values are adaptively and objectively set on these KPCs according to importance (P-WKPCA). Moreover, with the relationship between the previous and the current status considered, the status of the process is finally evaluated within a moving window to improve further the monitoring performance.

The rest of this article is structured as follows. First, the KPCA model used in process monitoring and KDE probability estimation are briefly reviewed, followed by a motivational example that illustrates the submerged useful information problem. Second, PM-WKPCA monitoring is proposed and some details are presented. In Section 4, the proposed monitoring method is tested in a numerical nonlinear process, the continuous stirred tank reactor (CSTR) process, and the Tennessee Eastman (TE) process. The monitoring results and several comparisons with conventional PCA, KPCA, and WPCA are presented. Finally, the conclusions in this study are given in Section 5.

2. Preliminaries

2.1. Kernel principal component analysis

In KPCA, observations are nonlinearly mapped into a high-dimensional feature space F and then linear PCA is employed to extract the nonlinear correlation between the variables [18,19,22]. Let the normalized training set be $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N \in R^m$ with N observations consisting of m measured process variables. The feature space is constructed by the nonlinear mapping: $R^m \rightarrow^{\boldsymbol{\Phi}(\bullet)} C^F$, where $\boldsymbol{\Phi}(\bullet)$ is the nonlinear mapping function [19,22]. The covariance matrix in the feature space F is calculated as

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

where $\sum_{j=1}^{N} \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right) = 0$ is assumed. The kernel principal component can be obtained by the eigenvalue problem [18,19,22]

$$\lambda \boldsymbol{v} = \boldsymbol{C}^F \boldsymbol{v} = \frac{1}{N}\sum_{j=1}^{N} \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)^T, \boldsymbol{v} \right\rangle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\lambda \boldsymbol{v} \tag{2}$$

where λ and $\mathbf{v}$ denote the eigenvalue and the eigenvector of the covariance matrix $\mathbf{C}^F$, respectively, and $\langle \boldsymbol{x},\boldsymbol{y} \rangle$ denotes the dot product between $\mathbf{x}$ and $\mathbf{y}$. This implies that all solutions $\mathbf{v}$ with $\lambda \neq 0$ must lie in the span of $\boldsymbol{\Phi}(\boldsymbol{x}_1),\ldots,\boldsymbol{\Phi}(\boldsymbol{x}_N)$. Then, $\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v}$ is equivalent to $\lambda \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{v} \rangle = \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{C}^F \boldsymbol{v} \rangle (k = 1,\ldots,N)$. There exist coefficients α_i $(i = 1,\ldots,N)$ such that $\boldsymbol{v} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\Phi}(\boldsymbol{x}_i)$. Rewriting Eq. (2), we obtain the following:

$$\lambda \sum_{i=1}^{N} \alpha_i \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{\Phi}(\boldsymbol{x}_i) \rangle = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left\langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \sum_{j=1}^{N} \left(\boldsymbol{x}_j\right) \right\rangle \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{\Phi}(\boldsymbol{x}_i) \right\rangle \quad (k = 1,\ldots,N). \tag{3}$$

Note that the eigenvalue problem in Eq. (3) only involves dot products of mapped shape vectors in the feature space. We define an $N \times N$ matrix $\mathbf{K}$ by $[\boldsymbol{K}]_{ij} = K_{ij} = \langle \boldsymbol{\Phi}(\boldsymbol{x}_i), \boldsymbol{\Phi}(\boldsymbol{x}_j) \rangle$. The left-hand side of Eq. (3) can then be expressed as [18,19,22]

$$\lambda \sum_{i=1}^{N} \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{\Phi}(\boldsymbol{x}_i) \rangle = \lambda \sum_{i=1}^{N} \alpha_i K_{ki} \quad (k = 1,\ldots,N) \tag{4}$$

and the right-hand side of Eq. (3) can be expressed as

$$\frac{1}{N}\sum_{i=1}^{N} \alpha_i \left\langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \sum_{j=1}^{N} \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right) \right\rangle \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{\Phi}(\boldsymbol{x}_i) \right\rangle = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \sum_{j=1}^{N} K_{kj} K_{ji} \quad (k = 1,\ldots,N). \tag{5}$$

The principal component $\mathbf{t}$ of a test vector $\mathbf{x}$ is then extracted by projecting $\boldsymbol{\Phi}(\boldsymbol{x})$ onto the eigenvectors v_k in F, as follows:

$$\boldsymbol{t}_k = \langle \boldsymbol{v}_k, \boldsymbol{\Phi}(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \boldsymbol{\Phi}(\boldsymbol{x}_i), \boldsymbol{\Phi}(\boldsymbol{x}) \rangle \quad (k = 1,\ldots,p) \tag{6}$$

where p is the number of the kernel principal components retained, which is empirically determined as the smallest number of the ordered eigenvalues whose cumulative sum is above 85%.

In KPCA monitoring, the T^2 and Q statistics are constructed and monitored based on the assumption that the training data have a multivariate normal distribution in the feature space [18]. The T^2 statistic is the sum of normalized squared scores, expressed as [9,19]

$$T^2 = \left[t_1,\ldots,t_p\right] \boldsymbol{\Lambda}^{-1} \left[t_1,\ldots,t_p\right]^T \tag{7}$$

where t_k is obtained from Eq. (6), and $\boldsymbol{\Lambda}^{-1}$ is the diagonal matrix of the inverse of the eigenvalues with the retained PCs.

The Q statistic is calculated by

$$Q = \left\| \boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}_p(\boldsymbol{x}) \right\|^2 = \left\| \overline{\boldsymbol{\Phi}}(\boldsymbol{x}) - \overline{\boldsymbol{\Phi}_p}(\boldsymbol{x}) \right\|^2 = \sum_{j=1}^{n} t_j^2 - \sum_{j=1}^{p} t_j^2 \tag{8}$$

where n is the number of nonzero eigenvalues among the total N eigenvalues. More details on the two statistic parameters are presented in the corresponding references [8,18,19].