Chemometrics and Intelligent Laboratory Systems 155 (2016) 151–159

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

# Performance-driven optimal design of distributed monitoring for large-scale nonlinear processes

Qingchao Jiang [a], Juan Li [b], Xuefeng Yan [a,*]

[a] 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
[b] College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao 266109, PR China

ARTICLE INFO

*Article history:*
Received 28 December 2015
Received in revised form 19 March 2016
Accepted 21 April 2016
Available online 23 April 2016

*Keywords:*
Distributed monitoring
Large-scale nonlinear process
Bayesian fault diagnosis
Kernel principal component analysis

ABSTRACT

Centralized monitoring generally involves all measured variables in one model. However, the existence of variables without beneficial information may cause redundancy in the monitoring and degrade monitoring performance. This paper proposes a performance-driven distributed monitoring scheme that incorporates kernel principal analysis (KPCA) and Bayesian diagnosis system for large-scale nonlinear processes. First, a stochastic optimization method is utilized to select a subset of variables that provide the best possible performance for each fault and to decompose the process into several sub-blocks. Second, a KPCA model is established in each block to deal with nonlinearity and generate fault signature evidence. Finally, a Bayesian fault diagnosis system is established to identify the fault status of the entire process. Considering the significant calculation amount in Bayesian diagnosis, optimal evidence source selection is performed to reduce the redundancy. Case studies on the Tennessee Eastman benchmark process and a continuous stirred tank reactor process demonstrate the efficiency of the proposed scheme.



## 1. Introduction

Modern chemical processes are usually characterized by large-scale and complex correlations, and monitoring such plant-wide processes has become an important issue [1–6]. Data-based, especially multivariate statistical process monitoring (MSPM), methods are gaining increasing attention because of advances in data gathering, transmitting, and processing techniques [7–10]. Numerous MSPM methods, such as principal component analysis (PCA) [11–14], partial least square [15–17], independent component analysis [18–20], and Fisher discriminate analysis [21–23], have been proposed to address various process characteristics, and the efficiencies of these methods have been proven. However, some problems still need to be addressed, such as (i) incorporating MSPM to a large-scale process; (ii) dealing with process nonlinearity; and (iii) performing fault diagnosis.

Nowadays, the number of measured variables in a plant-wide process is usually large. Incorporating all measured variables in one MSPM monitoring model is not appropriate for a large-scale process and may degrade monitoring performance [24,25]. To reduce process complexity and monitoring redundancy, a multi-block or distributed monitoring scheme that divides the process into several local units can be employed [3,5,15,25–27], in which process decomposition is a key step. The influence of process decomposition on monitoring performance has been analyzed in [25], and the detectability of a fault has been found to rely on both variable relations and fault characteristics. Therefore, in data-based process decomposition, both variable correlations and fault information should be taken into account. When fault information is unavailable, the main concern is the correlations among variables [5,26]; when fault information, i.e., data or model, is available, a performance-driven process decomposition method can be employed [25]. Ref. [25] established the theoretical foundation for data-driven distributed monitoring and addressed the fault detection issue for linear processes. Following the work in [25], this paper aims to extend the distributed monitoring scheme to large-scale nonlinear processes and to address both fault detection and diagnosis issues.

Aside from its large scale, a plant-wide process is usually characterized by nonlinearity; several approaches have been developed for dealing with this nonlinearity. For instance, an auto-associative neural network-based nonlinear PCA approach is proposed by Kramer [28], and a nonlinear PCA based on principal curves and neural networks is developed by Dong and McAvoy [29]. To avoid nonlinear optimization in neural network-based nonlinear PCA methods, kernel PCA (KPCA) has been proposed and extended intensively [30,31]. To employ KPCA for monitoring large-scale processes, a mutual information-based multiblock monitoring scheme has been proposed in [2]. This scheme takes both linear correlations and nonlinear relations among variables into account during process decomposition; however, without considering fault information, the optimal performance cannot be guaranteed.

Another issue in monitoring large-scale nonlinear processes is fault diagnosis. In practice, some faults occur constantly or periodically, and the data of these faults are stored in a historical database. The purpose

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

http://dx.doi.org/10.1016/j.chemolab.2016.04.012
0169-7439/

of fault diagnosis is to assign the current online sample to the most related historical fault class, which is essential for counteracting or eliminating the fault [21,23,32]. Rather than improving the diagnosis performance from the feature extraction aspect, Bayesian fault diagnosis makes decisions on the fault status in a probabilistic manner. A Bayesian diagnosis system that incorporates prior process knowledge is developed in [33], and a Bayesian control loop diagnosis framework is established by Huang [34]. Following these two works, several studies have been conducted, such as Refs. [35–38]. More recently, an optimal principal component-based Bayesian fault diagnosis system has been established in [39]. The work analyzed the influence of principal component (PC) selection on the diagnosis performance, and it suggested selecting the most efficient PCs to generate fault signature evidence. Following the work in [39], this paper will introduce how an optimal Bayesian diagnosis system can be designed in the distributed monitoring problem of a large-scale nonlinear process.

The remainder of this article is structured as follows: First, the basics of KPCA fault detection and Bayesian diagnosis are introduced briefly, and some notations are defined. Second, the performance-driven optimal design of distributed monitoring scheme that incorporates KPCA and Bayesian diagnosis system is presented in detail. Then, the proposed distributed monitoring scheme is applied on the Tennessee Eastman (TE) benchmark process and a simulated continuous stirred tank reactor (CSTR) process. Finally, conclusions are drawn.

## 2. Preliminaries

The basics of KPCA fault detection and Bayesian fault diagnosis are introduced briefly. Some notations are defined.

### 2.1. KPCA-based fault detection

KPCA is one of the most widely used techniques for dealing with the nonlinear process monitoring problem. Rather than decoupling variable correlations in a nonlinear input space, KPCA finds a computationally tractable solution through nonlinear mapping from the input space to the feature space [30]. Let $\boldsymbol{x}_k \in R^m, k=1,\ldots,N$ ($m$ is the number of measured variables, and $N$ is the number of samples) denote a set of zero-mean data. Given a nonlinear function $\Phi(\cdot)$, the covariance in a feature space $F$ can be expressed as [30,31]

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

where $\sum_{k=1}^{N}\Phi(\boldsymbol{x}_k)=0$ is assumed. In the feature space, the covariance matrix can be diagonalized by solving the eigenvalue problem as [30,31]

$$\lambda \boldsymbol{v} = \boldsymbol{C}^F \boldsymbol{v} \tag{2}$$

where $\lambda \geq 0$ are eigenvalues, and $\boldsymbol{v}$ are eigenvectors. According to Eq. (1), the right side of Eq. (2) becomes [30,31]

$$\boldsymbol{C}^F \boldsymbol{v} = \left(\frac{1}{N}\sum_{j=1}^{N} \Phi(\boldsymbol{x}_j)\Phi(\boldsymbol{x}_j)^T\right)\boldsymbol{v} = \frac{1}{N}\sum_{j=1}^{N}\langle \Phi(\boldsymbol{x}_j), \boldsymbol{v}\rangle \Phi(\boldsymbol{x}_j) \tag{3}$$

where $\langle\cdot,\cdot\rangle$ denotes the dot product. This finding implies that coefficients $\alpha_i (i=1,\ldots,N)$ that satisfy $\boldsymbol{v} = \sum_{i=1}^{N}\alpha_i \Phi(\boldsymbol{x}_i)$ exist because all solutions $\boldsymbol{v}$ with $\lambda \neq 0$ must lie in the span of $\Phi(\boldsymbol{x}_1),\ldots,\Phi(\boldsymbol{x}_N)$. Then, Eq. (2) can be expressed as [30,31]

$$\lambda\sum_{i=1}^{N}\alpha_i\langle \Phi(\boldsymbol{x}_k), \Phi(\boldsymbol{x}_i)\rangle = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\left\langle \Phi(\boldsymbol{x}_k), \sum_{j=1}^{N}\Phi(\boldsymbol{x}_j)\right\rangle\langle \Phi(\boldsymbol{x}_j), \Phi(\boldsymbol{x}_i)\rangle \tag{4}$$

for all $k=1,\ldots,N$. Introducing an $N\times N$ kernel matrix $\boldsymbol{K}$ by $[\boldsymbol{K}]_{ij}=K_{ij}=\langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j)\rangle$ ensures that the computing of $\Phi(\cdot)$ can be avoided and Eq. (4) can be expressed as [30,31]

$$\lambda N\boldsymbol{K}\boldsymbol{\alpha} = \boldsymbol{K}^2\boldsymbol{\alpha} \tag{5}$$

where $\boldsymbol{\alpha}=[\alpha_1,\ldots,\alpha_N]^T$. After some derivations, the PCs of a new sample $\boldsymbol{x}$ can be extracted as [30,31]

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

where $k=1,\ldots,p$, and $p$ is the number of retained PCs. Two statistics are constructed for monitoring the dominant space and residual space, respectively, as [30,31]

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

and

$$SPE = \left\|\Phi(x) - \hat{\Phi}_p(x)\right\|^2 = \sum_{j=1}^{n} t_j^2 - \sum_{j=1}^{p} t_j^2 \leq CL_{SPE} \tag{8}$$

where $n$ is the number of nonzero eigenvalues. The thresholds of the statistics can be obtained based on Gaussian assumption or determined by kernel density estimation. Here, Gaussian assumption is employed, and details on the determination of the thresholds are given in [30,31].

### 2.2. Bayesian fault diagnosis

In a physical process, some monitors indicate the statuses of process components, local units, or physical quantities. Occurrence of a fault will generally cause some changes to the process, and these changes will be reflected in the monitor readings. Bayesian diagnosis is used to identify the underlying fault status according to the obtained monitor readings and assign the current sample to the most related historical fault class. In using Bayesian diagnosis, the following notations should be defined [37].

Fault status $f$: A fault status $f$ denotes the underlying status of a process; for example, a fault status can be the normal operation state, a ramp change in temperature, or a disturbance in pressure. An $f$ can take different values in accordance with fault statuses, and the set of all $G$ fault statuses can be denoted as $f=\{f_1,\ldots,f_G\}$.

Evidence $\boldsymbol{E}$: Evidence $\boldsymbol{E}$ functions as the input for the Bayesian diagnosis system and is composed of a set of discrete values that indicate the states of the considered physical properties that come from the corresponding monitor readings. Each monitor reading can be regarded as a source in the evidence, and an evidence set with $r$ sources can be denoted as $E=\{\pi_1,\pi_2,\ldots,\pi_r\}$, where $\pi_i$ is the $i$-th source that has $q_i$ discrete values. The evidence set with all possible evidence values can be denoted as $\boldsymbol{\varepsilon}=\{e_1,e_2,\ldots,e_K\}$, where $\boldsymbol{e}_j \in \boldsymbol{\varepsilon}$ is a specific evidence and $K=\prod_{i=1}^{r} q_i$.

Historical evidence data $\boldsymbol{D}$: The historical training evidence data $\boldsymbol{D}$ refers to the evidence that come from each fault status collected in the process history and can be denoted as $\boldsymbol{D}=\{\boldsymbol{d}^1,\boldsymbol{d}^2,\ldots,\boldsymbol{d}^{N_D}\}$, where $N_D$ is the number of samples. The sample $\boldsymbol{d}^i$ taken at time $i$ consists of an evidence vector $\boldsymbol{e}^i$ and the fault status $f^i$: $\boldsymbol{d}^i=\{\boldsymbol{e}^i,f^i\}$. The samples in the historical dataset are assumed to be independent.

Bayesian solution: Bayesian diagnosis is used to identify the underlying fault status on the basis of the obtained evidence and historical evidence data. Results can be achieved by calculating the posterior