Journal of Process Control 68 (2018) 129–144

Contents lists available at ScienceDirect

Journal of Process Control

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

A novel hybrid of auto-associative kernel regression and dynamic independent component analysis for fault detection in nonlinear multimode processes

Jungwon Yu[a], Jaeyeong Yoo[b], Jaeyel Jang[c], June Ho Park[a], Sungshin Kim[a,*]

[a] Department of Electrical and Computer Engineering, Pusan National University, Busan, South Korea
[b] CTO, XEONET Co., Ltd, Seongnam, South Korea
[c] Technology & Information Department, Technical Solution Center, Korea East-West Power Co., Ltd., Dangjin, South Korea

ARTICLE INFO

Article history:
Received 2 November 2017
Received in revised form 12 March 2018
Accepted 12 May 2018

Keywords:
Auto-associative kernel regression
Dynamic independent component analysis
Fault detection
Nonlinear multimode processes

ABSTRACT

With modern industrial processes becoming larger and more complex, we should consider their nonlinear and multimode characteristics carefully for accurate process monitoring and fault detection. In this paper, a novel hybrid of two data-driven techniques—auto-associative kernel regression (AAKR) and dynamic independent component analysis (DICA)—is proposed for fault detection of nonlinear multimode processes. AAKR is a nonparametric multivariate technique; it can effectively deal with nonlinearity and multimodality of target systems by real-time local modeling in accordance with query vectors. Residuals obtained from AAKR usually deviate from Gaussian distribution (i.e., they are non-Gaussian), and there exist auto- and cross-correlations between them. The proposed method detects process faults by applying DICA to the residuals; DICA can capture useful statistical information hidden in the residuals. The validity and effectiveness of the proposed method are illustrated through three popular benchmark problems such as a three-variable multimodal process, a three-variable nonlinear process, and Tennessee Eastman process; the proposed method is also compared with several comparison methods The experimental results demonstrate the superiority of the proposed method, which achieves the best detection rates with reasonable false alarm rates.

1. Introduction

Modern industrial processes, such as power plants and manufacturing and chemical processes, have become increasingly larger and more complex; they operate with a large number of process variables. Recent advancements in measurement and communications technologies and implementation of distributed control systems significantly improve the quality and quantity of observed sensor data; they also enable us to consistently collect and manage the large volumes of operational data. As a consequence, these developments spark great interest in data-based techniques to minimize plant downtime and optimize process operations by analyzing the massive amounts of collected data.

Accurate and timely detection and diagnosis of possible faults enhance the availability, safety, and reliability of target systems and permit cost-effective operations. The faults mentioned previously are defined as unpermitted deviations of at least one characteristic property or variable of the target systems [1]. In the early stages of a fault, the effects on system performance may be negligible. However, if faults are left unattended without proper corrective actions, they may lead to severe performance degradation and can eventually cause system failures. Properly designed fault detection systems can provide the exact process operating conditions to operators and maintenance personnel, and can help them take appropriate remedial actions to remove potential abnormal behaviors occurring in target processes; they allow planned operations to be carried out successfully, and increase the productivity of process operations.

Principal component analysis (PCA) and independent component analysis (ICA) [6,7] are the most popular multivariate statistical techniques; they have been widely employed in the fields of process monitoring and fault detection [2–5]. Detecting possible process faults through traditional PCA and ICA, we only examine

* Corresponding author at: Engineering Bldg #10, 10314, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241, South Korea.
E-mail addresses: garden0312@pusan.ac.kr (J. Yu), joeyoo@xeonet.co.kr (J. Yoo), zanggre250@ewp.co.kr (J. Jang), parkjh@pusan.ac.kr (J.H. Park), sskim@pusan.ac.kr (S. Kim).

https://doi.org/10.1016/j.jprocont.2018.05.004
0959-1524/

static properties, but cannot catch dynamic behaviors of target systems. To overcome this limitation, dynamic PCA and ICA (DPCA and DICA) [8–10] were proposed. DPCA and DICA can capture not only the cross-correlations but also auto-correlations on each measured variable between its present value and those of the past; thus, they may be effective approaches to dynamic process monitoring.

The methods mentioned so far (i.e., PCA, ICA, DPCA, and DICA) are unsuitable for nonlinear process monitoring because they explicitly assume linearity between process variables. Kernel PCA and ICA (KPCA and KICA) [11–14], combining kernels with traditional PCA and ICA, have been successfully used to handle the nonlinearity of process data. Compared with previously suggested methods [15–17], in KPCA and KICA, the number of variables in reduced latent space does not need to be fixed in advance, and nonlinear optimization procedures are also not needed. However, the bigger the number of samples used for training, the longer the training time takes. In addition, it is difficult to determine proper kernel functions and their parameters, and to carry out contribution analysis to investigate root sources and/or causes of detected abnormal events.

In addition to the nonlinearity between monitored variables mentioned above, properly addressing multimodality of target systems is another important issue. Several methods have been proposed to monitor the condition of multimode processes, such as Gaussian mixture models (GMMs), hidden Markov models (HMMs), PCA mixture model, and adjoined ICA-PCA model [18–21,32,33]. Although these methods can tackle the multimodality of target processes successfully, model parameters obtained by expectation-maximization algorithm may often be trapped in a local optimum depending on initial starting points, and the number of modes (i.e., the number of clusters) should be predefined before searching for the parameters.

In this paper, a novel hybrid of auto-associative kernel regression (AAKR) [22–24] and DICA is proposed for fault detection of nonlinear and multiple-mode processes. AAKR belonging to lazy learning is a nonparametric multivariate technique to predict new query vectors by updating local models online; it does not need to be concerned about target data properties (i.e., linear or nonlinear; unimodal or multimodal) beforehand. In standard AAKR, after obtaining predicted vectors, residual vectors (i.e., error vectors) and squared prediction error (SPE) used as a detection index are calculated for fault detection.

Components of the residual vectors generated by AAKR, in general, may not follow Gaussian distribution accurately. Furthermore, there may exist cross-correlations between the residuals, and serial correlations in each residual component. The proposed method employs DICA to analyze the residual components obtained by AAKR; DICA can efficiently handle both the non-Gaussianity of the residuals and statistical relationships between them. From the hidden variables extracted from the residuals by DICA, we calculate three detection indices, such as I_d^2,I_e^2, and SPE statistics (also used in [4,10]), and then perform fault detection; kernel density estimation (KDE) is used to determine their upper threshold values.

The proposed method is based on the following key assumptions. First, we assume that monitored variables of multivariate data collected from target processes may have nonlinear relationships, and may follow not unimodal but multimodal distributions; if conventional PCA and ICA are directly applied to the process data, their performance may be degraded. Nonparametric AAKR can generate residual vectors where the nonlinearity and multimodality are properly removed; therefore, it is expected that the proposed method can improve fault detection performance. Second, it is assumed that the generated residuals may be correlated to each other, and there may exist serial correlations in each residual component; SPE statistic (see Eq. (5)) cannot consider the statistical properties hidden in the residuals. Third, the residuals are assumed to follow non-Gaussian distributions. Residual analysis via DICA can deal with these statistical properties, and enhance the performance of fault detection.

To verify the performance, the proposed and comparison methods are applied to three popular benchmark problems: a three-variable multimodal process, a three-variable nonlinear process, and Tennessee Eastman (TE) process. Experimental results show that the proposed method tackles the nonlinearity and multimodality of target data successfully, and achieves better fault detection rates (FDRs) than comparison methods; it is also confirmed that the performance of the proposed method is considerably enhanced by DICA that can handle non-Gaussianity plus auto- and cross-correlations of the residuals.

The remainder of this paper is organized as follows. Section 2 explains standard AAKR and fault detection based on it. Section 3 describes ICA and DICA for analyzing residual components generated by AAKR. Section 4 briefly outlines the proposed method (a novel hybrid of AAKR and DICA). Section 5 presents the experimental results and discussion, and finally, we give our conclusions and suggest future works in Section 6.

2. AAKR-based fault detection

In this section, AAKR-based fault detection is explained in detail; the followings are described with reference to the contents of Refs. [22,25].

2.1. Auto-associative kernel regression

From now on, scalars and vectors are written in italics and bold lowercase, respectively, and matrices are written in bold capitals. Let $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^n]^T \in \Re^{n\times m}$be a training data matrix composed of n data vectors $\mathbf{x}^i$, $i = 1, \ldots, n$, collected from a target system where each data vector consists of m observed process variables, i.e., $\mathbf{x}^i \in \Re^m$. The data vectors that constitute the matrix should encompass the whole operational range of normal target system behaviors. There are various distance functions to measure the similarities between the data vectors and the current query vector, such as Euclidean, Manhattan, Chebychev, and Mahalanobis. In this study, among the widely used distance functions, we employ Euclidean or Mahalanobis distance functions, defined as

$$d_E^i(\boldsymbol{x}^i, \boldsymbol{x}_{\text{new}}) = \sqrt{(\boldsymbol{x}^i - \boldsymbol{x}_{\text{new}})^T(\boldsymbol{x}^i - \boldsymbol{x}_{\text{new}})}\ , i = 1, \ldots, n \tag{1}$$

$$d_M^i(\boldsymbol{x}^i, \boldsymbol{x}_{\text{new}}) = \sqrt{(\boldsymbol{x}^i - \boldsymbol{x}_{\text{new}})^T \boldsymbol{S}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}^i - \boldsymbol{x}_{\text{new}})}\ , i = 1, \ldots, n \tag{2}$$

where $\mathbf{x}_{\text{new}}$ is a new query vector, $\mathbf{x}^i$ is the ith training vector stored in memory, and $\boldsymbol{S}_{\boldsymbol{x}} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X}$is a covariance matrix calculated from the data matrix $\mathbf{X}$. Depending on the similarities measured by the distance functions, weights for each training data vector are generated by a weighting function. In this paper, among several commonly used weighting functions (e.g., Gaussian, triangular, exponential, quadratic, and tricube), Gaussian weighting function is used to assign the weights as follows:

$$K_h(d^i) = \frac{1}{\sqrt{2\pi}h} \exp\left[-\frac{(d^i)^2}{2h^2}\right], i = 1, \ldots, n \tag{3}$$

where h is a bandwidth parameter in the weighting function, and d^i is the distance function value between the ith training data vector and a new query vector. The shape of Eq. (3) is identical to those of probability density function for GMMs [18] and covariance function in Gaussian process [37]; different from them, Gaussian weighting function in AAKR assigns weights to memory vectors according to their similarities with query vectors.