ISA Transactions ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

ISA Transactions

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

Research article

Double-layer ensemble monitoring of non-gaussian processes using modified independent component analysis

Chudong Tong*, Ting Lan, Xuhua Shi

Faculty of Electrical Engineering & Computer Science, Ningbo University, Ningbo 315211, PR China

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form
21 December 2016
Accepted 7 February 2017

Keywords:
Modified independent component analysis
Fault detection
Ensemble learning

ABSTRACT

The modified independent component analysis (MICA) was proposed mainly to obtain a consistent solution that cannot be ensured in the original ICA algorithm and has been widely investigated in multivariate statistical process monitoring (MSPM). Within the MICA-based non-Gaussian process monitoring circle, there are two main problems, i.e., the selection of a proper non-quadratic function for measuring non-Gaussianity and the determination of dominant ICs for constructing latent subspace, have not been well attempted so far. Given that the MICA method as well as other MSPM approaches are usually implemented in an unsupervised manner, the two problems are always solved by some empirical criteria without respect to enhancing fault detectability. The current work aims to address the challenging issues involved in the MICA-based approach and propose a double-layer ensemble monitoring method based on MICA (abbreviated as DEMICA) for non-Gaussian processes. Instead of proposing an approach for selecting a proper non-quadratic function and determining the dominant ICs, the DEMICA method combines all possible base MICA models developed with different non-quadratic functions and different sets of dominant ICs into an ensemble, and a double-layer Bayesian inference is formulated as a decision fusion method to form a unique monitoring index for online fault detection. The effectiveness of the proposed approach is then validated on two systems, and the achieved results clearly demonstrate its superior proficiency.

1. Introduction

Modern industrial plants have been witnessing a rapid development of distributed computer-aided systems and sensor technologies as well as operator support systems through data-driven process monitoring systems [1–3]. Among diverse process monitoring techniques, multivariate statistical process monitoring (MSPM) provides a data-based framework for conducting such activities without accurate first-principal models, and has become a popular subject of active research in recent years [4–6]. Two fundamental MSPM methods, principal component analysis (PCA) and partial least squares (PLS), have been intensively investigated because of the capability to deal with a large number of strongly cross-correlated variables. However, the proficiency of identifying faults from data for the PCA/PLS-based methods can be deteriorated by an assumption that the process data approximately follow a multivariate Gaussian distribution [7–9]. Fortunately, independent component analysis (ICA), which utilizes higher-order statistics to glean more information and makes distribution of the projected data as independent as possible, has found wider applications in non-Gaussian process monitoring as we as other research fields such as biomedical system, signal processing, and so on [8,10–15].

An examination of the existing literature on ICA shows that the FastICA iterative algorithm proposed by Hyvärinen and Oja [16] is used as a "default" method for model construction because it can greatly reduce the computation time. The FastICA algorithm has been intensively adopted for a functional data processing tool, especially in the area of blind source separation [17–20]. However, the original FastICA algorithm has some drawbacks from practical viewpoint. First, inconsistent solutions would be given because of its random initialization procedure. Second, the importance of extracted independent components (ICs) is not ordered. To attempt these issues, Lee et al. [9] proposed a modified ICA (MICA) algorithm that extracts a few dominant ICs, determines the order of ICs, and gives a consistent solution. The basic idea is to first estimate variance of dominant ICs and the directions using PCA and then to perform the FastICA algorithm to update the dominant ICs while maintaining the variance. Furthermore, Zhang and Zhang [8] adopted the particle swarm optimization (PSO) method for ICs extraction, and the order of ICs is then determined according to the role of resumption of the original data. Although the

* Corresponding authors.
E-mail address: tongchudong@nbu.edu.cn (C. Tong).

http://dx.doi.org/10.1016/j.isatra.2017.02.003
0019-0578/

Please cite this article as: Tong C, et al. Double-layer ensemble monitoring of non-gaussian processes using modified independent component analysis. ISA Transactions (2017), http://dx.doi.org/10.1016/j.isatra.2017.02.003

PSO-based ICA algorithm would reduce the risk of obtaining local minimum solution, it significantly increases the computation time compared with the FastICA algorithm. Shen et al. [21] proposed a variational Bayesian ICA algorithm and demonstrated its improved stability and separating ability for blind source separation. Recently, Ge and Song [22] developed a performance-driven ensemble learning ICA model for non-Gaussian process monitoring. The ensemble learning approach is used to improve the stability of the FastICA algorithm, and the determination of dominant ICs is realized by a performance-driven method with reference abnormal datasets involved. Additionally, Tong et al. [23] adopted the ensemble learning technique to solve the problem derived from the determination logic of dominant ICs for constructing latent subspace. The fault detectability of their method for non-Gaussian processes is greatly enhanced. The utilization of ensemble learning strategy has become quite popular in the field of MSPM over the last years [23–25]. This is mainly because the ensemble learning aims to combine multiple solutions into an ensemble one, which is always shown to have a better performance than the outcome of individual solutions.

Nevertheless, it should be stressed that the previously mentioned ICA methods all involve a measure of non-Gaussianity so as to reflect the statistically independence of ICs. The negentropy based on the information theoretic quantity of differential entropy is usually served as a good measure of non-Gaussianity of a random variable. Hyvärinen and Oja [16] introduced a flexible and reliable approximation of negentropy through using a proper non-quadratic function, and suggested three formulations for the non-quadratic functions. An additional problem is thus raised from the selection of a proper non-quadratic function. Given that the ICA-based approaches as well as other MSPM methods is unsupervised, which means that only a dataset collected under normal operating condition is needed, the selection of a proper non-quadratic function becomes inaccessible. Meanwhile, the available samples from all possible faulty conditions is highly limited in practice, the determination of a proper non-quadratic function can thus not be solved directly through testing different reference abnormal datasets empirically. Therefore, although the two main challenges mentioned previously are well addressed, the selection of a proper non-quadratic function is an additional problem that still remains unsolved to date.

This paper aims to develop an efficient non-Gaussian process monitoring method that can well handle all these issues raised in the traditional ICA monitoring scheme. The MICA algorithm is borrowed here for modeling non-Gaussian process data since it can produce a consistent result. First, the iterative procedures in the MICA algorithm are implemented three times in parallel, with three different non-quadratic functions involved. Second, within each MICA model, the dominant ICs are not only determined by their variance but also be determined by other available criteria that will be introduced in the next section. After these two modeling steps, a series of MICA monitoring models are thus resulted. Third, to obtain a unique monitoring decision for newly sampled measurements, a double-layer ensemble monitoring scheme based on Bayesian inference is developed. Benefiting from the superiority of ensemble learning trick, the proposed double-layer ensemble monitoring method based on MICA (abbreviated as DEMICA) takes all possibilities of base MICA models into account, the previously presented issues (*i.e.*, determination logic of dominant ICs and selection logic of non-quadratic functions) are no longer troublesome and the monitoring performance can always be enhanced.

2. MICA-based process monitoring

2.1. MICA algorithm

The first step of MICA method is to use PCA to extract all principal components (PCs) from process dataset $\mathbf{X} \in R^{m\times n}$:

$$\mathbf{T} = \mathbf{P}^{\mathrm{T}}\mathbf{X} \tag{1}$$

where m and n are the number of measured variables and samples, respectively. $\mathbf{T} \in R^{m\times n}$ consists of the extracted PCs, $\mathbf{P} \in R^{m\times m}$ contains the eigenvectors of covariance matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(n-1) = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}$, and $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$. In some cases, the last few eigenvalues in $\boldsymbol{\Lambda}$ are close to zero, they are suggested to be excluded, but it is important to retain as many eigenvalues as possible. The normalized PCs, given as

$$\mathbf{Z} = \boldsymbol{\Lambda}^{-1/2}\mathbf{T} = \boldsymbol{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}}\mathbf{X} = \mathbf{Q}\mathbf{X} \tag{2}$$

is then served as an initial estimate for ICs, where $\mathbf{Q} = \boldsymbol{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}}$.

The objective of MICA algorithm is to find a matrix $\mathbf{C} \in R^{m\times d}$ satisfying $\mathbf{C}^{\mathrm{T}}\mathbf{C} = \mathbf{D}$, whose form is such that the extracted, given as

$$\mathbf{S} = \mathbf{C}^{\mathrm{T}}\mathbf{Z} \tag{3}$$

become as independent of each other as possible, where $\mathbf{D} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_d\}$. The requirement $\mathbf{S}\mathbf{S}^{\mathrm{T}}/(n-1) = \mathbf{D}$ reflects that the variance of each IC in $\boldsymbol{S}$ is the same as that of the corresponding PC in PCA. Therefore, the importance of the ICs in MICA can be ordered according to their variance. By defining the normalized ICs as

$$\mathbf{S}_n = \mathbf{D}^{-1/2}\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{C}^{\mathrm{T}}\mathbf{Z} = \mathbf{C}_n{}^{\mathrm{T}}\mathbf{Z} \tag{4}$$

with $\mathbf{C}_n{}^{\mathrm{T}} = \mathbf{D}^{-1/2}\mathbf{C}^{\mathrm{T}}$ and $\mathbf{C}_n{}^{\mathrm{T}}\mathbf{C}_n = \mathbf{I}$, the essence of MICA is therefore to find the matrix $\mathbf{C}_n$. Once the $\mathbf{C}_n$ is found, the demixing matrix $\mathbf{W} \in R^{d\times m}$ and mixing matrix $\mathbf{A} \in R^{m\times d}$ for ICA model construction can be calculated as follows

$$\mathbf{W} = \mathbf{D}^{1/2}\mathbf{C}_n{}^{\mathrm{T}}\mathbf{Q} = \mathbf{D}^{1/2}\mathbf{C}_n{}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}} \tag{5}$$

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{C}_n\mathbf{D}^{-1/2} \tag{6}$$

where $\mathbf{W}\mathbf{A} = \mathbf{I}_d \in R^{d\times d}$. The extracted dominant ICs reveal the majority of information and represent a meaningful representation about the observed data $\boldsymbol{X}$. The relation between the modified ICs and PCs are:

$$\mathbf{S} = \mathbf{W}\mathbf{X} = \mathbf{D}^{1/2}\mathbf{C}_n{}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}}\mathbf{X} = \mathbf{D}^{1/2}\mathbf{C}_n{}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{T} \tag{7}$$

Given that the variance of each IC in $\boldsymbol{S}$ is the same as that of the corresponding PC in PCA, the number of retained ICs, d, can then be determined by some criteria that used in PCA, such as cumulative percent variance (CPV), cross-validation [26], and the variance of the reconstruction error [27]. If the original data follows an Gaussian distribution, no further dependency exists beyond the second-order moments of $\boldsymbol{X}$, and thus $\mathbf{C}_n$ converges to $[\mathbf{I}_d \vdots \mathbf{0}]$ and $\mathbf{S} = \mathbf{T}$. In this special case, the MICA converges to PCA. The detailed iterating procedures of MICA algorithm is provided in Lee et al. [9].

2.2. Fault detection based on MICA

The implementations of the monitoring statistics of the MICA are similar to those of the monitoring statistics of the PCA method. Usually, two statistics (*i.e.*, T^2 and Q) are computed in the MICA model. Since a total of d $(d << m)$ dominant ICs is retained to span a reduced feature space, the T^2 and Q statistics aimed for monitoring the variations in feature space and residual are formulated as follows

Please cite this article as: Tong C, et al. Double-layer ensemble monitoring of non-gaussian processes using modified independent component analysis. ISA Transactions (2017), http://dx.doi.org/10.1016/j.isatra.2017.02.003