Control Engineering Practice 58 (2017) 34–41

Contents lists available at ScienceDirect

Control Engineering Practice

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

Ensemble modified independent component analysis for enhanced non-Gaussian process monitoring

Chudong Tong*, Ting Lan, Xuhua Shi

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

ARTICLE INFO

Keywords:
Modified independent component analysis
Fault detection
Ensemble learning

ABSTRACT

As a multivariate statistical tool, the modified independent component analysis (MICA) has drawn considerable attention within the non-Gaussian process monitoring circle since it can solve two main problems in the original ICA method. Despite the diversity in applications, the determination logic for non-quadratic functions involved in the iterative procedures of MICA algorithm has always been empirical. Given that the MICA is an unsupervised modeling method, a direct rational study that can conclusively demonstrate which non-quadratic function is optimal for the general purpose of fault detection is inaccessible. The selection of non-quadratic functions is still a challenge that has rarely been attempted. Recognition of this issue and motivated by the superiority of ensemble learning strategy, a novel ensemble MICA (EMICA) modeling approach is presented for enhancing non-Gaussian process monitoring performance. Instead of focusing on a single non-quadratic function, the proposed method combines multiple base MICA models derived from different non-quadratic functions into an ensemble one, and the Bayesian inference is employed as a decision fusion method to form a unique monitoring index for fault detection. The enhanced fault detectability of the EMICA method is also illustrated on two industrial processes.

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, in particular, multivariate statistical process monitoring (MSPM) techniques in recent years (Ruiz-Cárcel, Cao, Mba, Lao, & Samuel, 2015; Yin, Li, Gao, & Kaynak, 2015). Not surprisingly, MSPM on the basis of two fundamental algorithms: principal component analysis (PCA) and partial least squares (PLS), has been receiving considerable attention as first-principle models of modern complex process systems are often costly to develop or practically inaccessible (Portnoy, Melendez, Pinzon, & Sanjuan, 2016; Yin, Ding, Xie, & Luo, 2014). However, the proficiency of identifying faults from data for the PCA/PLS-based methods can be deteriorated because they assume that the sampled data follows a multivariate Gaussian distribution approximately (Lee, Qin, & Lee, 2006; Lee, Yoo, & Lee, 2004; Zhang & Zhang, 2010). To handle the monitoring problem of non-Gaussian processes, independent component analysis (ICA), which can extract more useful information from non-Gaussian process data with the utilization of higher-order statistics, has been intensively investigated in the last few years (Fan, Qin, & Wang, 2014; Hsu, Chen, & Chen, 2010; Jiang, Wang, & Yan, 2015; Lee et al., 2006; Zhang & Zhang, 2010). In comparison to PCA, ICA not only de-correlates the data but also makes the projected data as independent as possible, and thus glean more essential features from observed measurements.

Among the diverse applications of the ICA-based non-Gaussian process monitoring, the FastICA iterative algorithm proposed by Hyvärinen (1999) is always employed as a "default" method for model construction because it can greatly reduce the computation time. However, the utilization of the FastICA algorithm has some drawbacks in practical applications. First, different solutions would be obtained because of its random initialization step, which might result in unstable monitoring models. Second, unlike PCA model which sorts the importance of the principal components (PCs) according to their variance, a proper ordering of the independent components (ICs) is still a well-known issue. To tackle these challenges, Lee et al. (2006) developed a modified ICA (MICA) algorithm that extracts a small number of ordered ICs and produces a consistent result as well. The basic idea is to first use the normalized PCs from PCA model as an initial estimate for ICs and then to perform the FastICA algorithm to update the dominant ICs while maintaining the variance. Furthermore, Zhang and Zhang (2010) adopted the particle swarm optimization method for estimating ICs, and the importance of ICs is then sorted by

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

http://dx.doi.org/10.1016/j.conengprac.2016.09.014
Received 7 April 2016; Received in revised form 16 August 2016; Accepted 24 September 2016
0967-0661/

the role of resumption of the original data. Although the PSO-based ICA algorithm would reduce the risk of obtaining local minimum solution, it increases the computation time compared with the FastICA algorithm. Recently, Ge and Song (2013) 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. The essence of the ensemble learning is to combine multiple solutions from different models into a unique one, which is expected to give a significantly better result than any outcomes of individual solutions. Benefiting from this superiority, the ensemble learning technique has become quite popular in the field of MSPM over the last several years (Ge & Song, 2014; Li & Yang, 2014; Tong, Palazoglu, & Yan, 2014).

Nevertheless, it should be stressed that all these ICA modeling methods mentioned above involve a measure of non-Gaussianity so as to reflect the statistically independence of ICs. The negentropy on the basis of the information theoretic quantity of differential entropy, is usually served as a good estimate of non-Gaussianity of a random variable. However, the calculation of negentropy requires an estimation of the probability density function, which sometimes is unobtainable. Fortunately, Hyvärinen (1999) formulated a feasible and reliable calculation of negentropy through using a proper non-quadratic function. Given that there are three suggestions for the non-quadratic function available in the literature, the stability of the ICA model cannot be ensured with different non-quadratic functions employed, and thus the resulted monitoring performance would also be affected. Generally, the modeling procedures in the ICA-based process monitoring method as well as other approaches in MSPM are unsupervised, which means that only a dataset sampled under normal operating condition is needed. Without respect to abnormal data, a proper selection of the non-quadratic functions is inaccessible. Meanwhile, the available samples from all possible faulty conditions is highly limited, a single empirically determination of the non-quadratic functions would lead to some specific faults undetected. From this viewpoint, a single non-quadratic function cannot be effective for all kinds of faults. Therefore, the selection of non-quadratic functions is a severe problem that remains unsolved.

Recognition of this issue motivates the current study, which integrates the ensemble learning strategy into the MICA algorithm. As mentioned previously, the MICA modeling method can address the two challenges existed in the original ICA iterative procedures. Additionally, the MICA only extracts a few dominant ICs instead of all ICs that needed for process monitoring, high computational load can thus be attenuated (Lee et al., 2006). With the involvement of ensemble learning strategy, the proposed ensemble MICA (EMICA) method combines multiple base MICA models resulted from different non-quadratic functions into an ensemble one through Bayesian inference based decision fusion (Ghosh, Ng, & Srinivasan, 2011). Since any of these non-quadratic functions could be useful for fault detection, a feasible solution is to take advantage of all of them, and then produce an ensemble result for enhanced non-Gaussian process monitoring. Unlike traditional MICA-based monitoring method, multiple base MICA monitoring models with different non-quadratic functions utilized are first developed, the Bayesian inference strategy is then employed for online fault alarm decision fusion, which generates an ensemble probabilistic index from multiple monitoring statistics.

2. MICA based process monitoring

2.1. MICA algorithm

The first step of MICA method is to use PCA to extract all available PCs from data $\mathbf{X} \in R^{m\times n}$:

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

where $\boldsymbol{X}$ contains n samples of m measured variables. $\mathbf{T} \in R^{m\times n}$ consists of the extracted PCs, $\mathbf{P} \in R^{m\times m}$ is composed of the eigenvectors of covariance matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(n-1) = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}$, and $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$. The last few elements in $\boldsymbol{\Lambda}$ are sometimes close to zero because of the collinearity existed in the measurements, they can be excluded. But it is highly suggested to include as many eigenvalues as possible. The extracted PCs are whitened as follows:

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

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

The objective of MICA algorithm is to update a matrix $\mathbf{C} \in R^{m\times d}$ satisfying $\mathbf{C}^{\mathrm{T}}\mathbf{C} = \mathbf{D}$ with a form such that the extracted components

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

become as independent of each other as possible, where $\mathbf{D} = diag\{\lambda_1, \lambda_2, \ldots, \lambda_d\}$. The requirement $\mathbf{S}\mathbf{S}^{\mathrm{T}}/(n-1) = \mathbf{D}$ makes the variance of each IC in $\boldsymbol{S}$ and the corresponding PC in PCA be the same Therefore, a proper ordering of the ICs in MICA can then be realized in accordance with their variance. The $\boldsymbol{S}$ can be normalized by

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

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 main task of MICA is thus reduced to find the matrix $\mathbf{C}_n$. The demixing matrix $\mathbf{W} \in R^{d\times m}$ and mixing matrix $\mathbf{A} \in R^{m\times d}$ are given as

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

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

where $\mathbf{W}\mathbf{A} = \mathbf{I}_d \in R^{d\times d}$. 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 (Valle, Li, & Qin, 1999; Wold, 1978), for example, cumulative percent variance (CPV). If the process data strictly follows an Gaussian distribution, $\mathbf{C}_n$ reduces to $[\mathbf{I}_d \vdots \mathbf{0}]$, which means $\mathbf{S} = \mathbf{T}$. Therefore, the PCA can be considered as a special case of the MICA, and the updating of $\mathbf{C}_n$ can be started from a fixed initialization, *i.e.*, $[\mathbf{I}_d \vdots \mathbf{0}]$.

In the MICA iterative procedures provided in the Appendix, the statistically independent requirement of ICs needs a measure of non-Gaussianity. Generally, the measure of non-Gaussianity is approximated by

$$J(y) = [E\{G(y)\} - E\{G(v)\}]^2 \tag{7}$$

where y is scaled to be of zero mean and unit variance, v is a Gaussian variable of zero mean and unit variance, and G is known as the non-quadratic function. Hyvärinen and Oja (2000) introduced three non-quadratic functions:

$$G_1(u) = \frac{1}{a_1}\log\cosh(a_1 u) \tag{8}$$

$$G_2(u) = \exp(-a_2 u^2/2) \tag{9}$$

$$G_3(u) = u^4 \tag{10}$$

where $1 \le a_1 \le 2$ and $a_2 \approx 1$. It has been empirically shown that G_1 is a good contrast function, and it is usually adopted for ICA model construction. However, as will be illustrated later, the function G_1 cannot be always good for improving fault detectability since the MICA is an unsupervised modeling method. The selection of G would highly influence the coming monitoring results. Without enough prior knowledge, the optimal determination of function G is still an open problem.

2.2. Process monitoring based on MICA algorithm

Similar to PCA-based fault detection method, the implementation of MICA for fault detection also depends on two statistics (*i.e.*, T^2 and