Automatica 96 (2018) 73–83

Contents lists available at ScienceDirect

Automatica

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

Process monitoring using a generalized probabilistic linear latent variable model[☆]

Rahul Raveendran, Hariprasad Kodamana [1], Biao Huang [*]

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada

ARTICLE INFO

Article history:
Received 28 June 2017
Received in revised form 6 January 2018
Accepted 26 May 2018

Keywords:
Process monitoring
Multivariate techniques
Probabilistic latent variable models
EM algorithm

ABSTRACT

This paper defines a generalized probabilistic linear latent variable model (GPLLVM) that under specific restrictions reduces to various probabilistic linear models used for process monitoring. For the defined model, we rigorously derive the monitoring statistics and their respective null distributions. Monitoring statistics of the defined model also reduce to the monitoring statistics of various probabilistic models when restricted with the corresponding conditions. The paper presents insightful equivalence between the classical multivariate techniques for process monitoring and their probabilistic counterparts, which is obtained by restricting the generalized model. We also provide an estimation approach based on the expectation maximization algorithm (EM) for GPLLVM. The results presented in the paper are verified using numerical simulation examples.

1. Introduction

Process monitoring ensures product quality, process safety and equipment reliability of a process. Typically in a process, numerous variables are measured or inferred and archived. Continuous monitoring of these variables will ensure desired product quality is achieved, the operation is within the safe envelope and the health parameters of the equipment are within the allowable bounds. In the past few decades, the academic and industrial communities have shown increasing interest in multivariate techniques for monitoring (Alcala & Qin, 2009; Chiang, Russell, & Braatz, 2000; Jiang, Yan, & Huang, 2016; MacGregor & Kourti, 1995; Qin, 2003; Wise & Gallagher, 1996; Yin, Ding, Haghani, Hao, & Zhang, 2012). Multivariate techniques are used to characterize the normal operation of the process from the archived historical data. This is further used as a reference for monitoring the measured variables online.

Most popular multivariate techniques used for monitoring include principal component analysis (PCA), factor analysis (FA), partial least squares (PLS) and canonical correlation analysis (CCA) among others (Li, Qin, & Zhou, 2010; Negiz & Çlinar, 1997; Russell, Chiang, & Braatz, 2000; Wise & Gallagher, 1996; Xia, Howell, & Thornhill, 2005). They provide a significant dimension reduction by mapping the high dimension variables onto a lower dimension latent space. As a result, a few latent variables are obtained to represent the measured high dimension variables by retaining the desired characteristics of the high dimension variables. For instance, PCA provides dimension reduction on the output variables of the process by retaining maximal variance in the data and CCA extracts maximally correlated lower dimension latent variables from the input variables and the output variables (Anderson, 2003). Further, control/monitoring statistics are derived from the lower dimensional latent variables and monitored based on the control limits derived from the null distributions of the statistics (Qin, 2003).

Classical latent variable techniques also have their probabilistic counterparts. For instance, probabilistic PCA (PPCA) (Tipping & Bishop, 1999b) and probabilistic CCA (PCCA) (Bach & Jordan, 2005) are the probabilistic counterparts of PCA and CCA, respectively. Advantages of these probabilistic versions include: They provide (1) complete distribution models for the dataset, (2) feasible frameworks to accommodate different distribution assumptions to the dataset with specific data characteristics, for instance, outliers (Chen, Martin, & Montague, 2009) and multi-modality (Tipping & Bishop, 1999a), and (3) frameworks to handle missing data (Ilin & Raiko, 2010). Some of the probabilistic counterparts mentioned above have also been considered for process monitoring (Chen et al., 2009; Chen & Sun, 2009; Ge & Song, 2010; Jiang, Huang, & Yan, 2016; Kim & Lee, 2003; Zhao, Li, Huang, Liu, & Ge, 2015). However, the existing literature does not bring

☆ This work is supported by Natural Sciences and Engineering Research Council of Canada (Grant NO. IRCPJ417793 - 15). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Michele Basseville under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: rraveend@ualberta.ca (R. Raveendran), hariprasad.k@iitb.ac.in, kodamana@ualberta.ca, kodamana@chemical.iitd.ac.in (H. Kodamana), biao.huang@ualberta.ca (B. Huang).
[1] Currently with Department of Chemical Engineering, Indian Institute of Technology Delhi, Delhi, India.

https://doi.org/10.1016/j.automatica.2018.06.029
0005-1098/

enough insights on their connection with the classical multivariate techniques in the context of process monitoring.

As we intend to investigate the connection between the probabilistic models and the classical techniques in the context of monitoring, we find that there lies an incentive for defining a GPLLVM that encompasses various probabilistic models. Instead of looking at monitoring based on several models in isolation, it helps us view the monitoring approaches in a single coherent framework. It allows developing the monitoring approaches just for the generalized model, which then can be simplified effortlessly to the special cases if desired. Hence, the following are the objectives of this work: (1) Define a GPLLVM for monitoring that subsumes several probabilistic counterparts of the classical multivariate techniques, (2) develop monitoring statistics for the general model, (3) bring out the insightful connection between the probabilistic counterparts and the classical techniques in the context of monitoring, and (4) present an approach based on the EM algorithm for estimating the maximum likelihood parameters of the GPLLVM. Also, as a part of this exercise, we flag some common issues related to the monitoring statistics in the existing literature for probabilistic latent variable model based monitoring.

The remainder of this paper is organized as follows: In Section 2, we present the preliminaries required for the paper. In Section 3, we introduce the proposed model for monitoring. In Section 4, we present the development of monitoring statistics from the model. In Section 5, we show the connection between the monitoring methods based on probabilistic latent variable models and the classical techniques. In Section 6, we show numerical simulations for verifying the presented results and in Section 7, we present the concluding remarks. Also, we provide an approach based on the EM algorithm for estimating the proposed model in Appendix A for completion.

Notations: Here we present the recurring and the commonly used notations in the paper. $\mathbb{R}$ represents the space of real numbers, I_P represents identity matrix of size $P \times P$, $E(.)$ and $Cov(.)$, represent the expectation and covariance, respectively, $diag(.)$ represents the operator that converts a vector into a diagonal matrix, $\mathcal{N}(\mu, \Sigma)$ represents multivariate normal distribution with mean μ and covariance Σ and superscript ^{T} represents the transpose operation. Other notations used in the paper are described when they are first introduced.

2. Preliminaries

In this section, we provide a brief review of the PCA and CCA based monitoring approaches that is necessary to appreciate some of the key results presented in this paper.

2.1. PCA based monitoring

Consider a system with the mean-centred outputs $Y \triangleq \{y_1, \dots, y_n \in \mathbb{R}^P, \dots, y_N\} \in \mathbb{R}^{P\times N}$ with sample covariance matrix $\tilde{\Sigma}_{yy} \succeq 0$, corresponding to the normal operation of the system. Using SVD/eigendecomposition, $\tilde{\Sigma}_{yy}$ is decomposed as the following,

$$\tilde{\Sigma}_{yy} = \eta_K \Lambda_K \eta_K^T + \eta_{\sim K} \Lambda_{\sim K} \eta_{\sim K}^T \tag{1}$$

where $\mathbb{R}^{K\times K} \ni \Lambda_K \succ 0$ is a diagonal matrix with K, $(K < P)$, principal eigenvalues of $\tilde{\Sigma}_{yy}$ as the diagonal elements and $\mathbb{R}^{(P-K)\times(P-K)} \ni \Lambda_{\sim K} \succeq 0$ is a diagonal matrix with $(P-K)$ minor eigenvalues of $\tilde{\Sigma}_{yy}$ as the diagonal elements. The matrices $\eta_K \in \mathbb{R}^{P\times K}$ and $\eta_{\sim K} \in \mathbb{R}^{P\times K}$ are composed of orthonormal eigenvectors as columns corresponding to the eigenvalues in Λ_K and $\Lambda_{\sim K}$, respectively. Then, the minor eigenvectors are discarded and using η_K, the data is projected onto the lower dimension latent space as the following,

$$S_K = \eta_K^T Y \tag{2}$$

where $S_K = \{s_1, \dots, s_n \in \mathbb{R}^K, \dots, s_N\} \in \mathbb{R}^{K\times N}$ are the lower dimension latent variables. The latent variables are linearly uncorrelated and their covariance is given by $\Lambda_K = diag\,(\lambda_1, \lambda_2, \dots, \lambda_K)$.

The PCA model developed from the normal data would be deployed to monitor the routine operation data. Commonly, two different statistics namely, (1) Hotelling's T^2 (Hotelling, 1947), and (2) Q (Jackson & Mudholkar, 1979) are monitored to check the conformity of the new data to the normal operation. T^2 is the normalized sum of squares of latent variables. For an observation y_n, the T^2 statistic is defined as the following,

$$T_n^2 = \|\Lambda_K^{\frac{1}{2}} s_n\|_2 = s_n^T \Lambda_K^{-1} s_n = y_n^T \eta_K \Lambda_K^{-1} \eta_K^T y_n \tag{3}$$

The Q statistic is the sum of squares of the residuals obtained with optimal least squares reconstruction. When the data is reconstructed from the lower dimensional latent variables, the optimal reconstruction for an observation in the least squares sense is given by,

$$\hat{y}_n = \eta_K s_{nK} = \eta_K \eta_K^T y_n \tag{4}$$

where $\hat{y}_n$ is the reconstruction of y_n given by PCA, the reconstruction residual r_n is given by,

$$r_n = (I_P - \eta_K \eta_K^T) y_n \tag{5}$$

and the Q statistic is defined as the following,

$$Q_n = \|r_n\|_2 = r_n^T r_n = y_n^T \left(I_P - \eta_K \eta_K^T\right) y_n \tag{6}$$

Remark 1. It can also be shown that PCA is an optimal solution of $\min_{\eta_K} \Sigma_{n=1}^N \|r_n\|_2$ s.t $\eta_K^T \eta_K = I_K$.

The null distribution of the T^2 statistic is a χ^2 distribution with K degrees of freedom. The Q statistic is reducible to a nonnegative sum of χ^2 random variables, and for its cumulative distribution function, several approximations are available in the literature (for a recent review, see Bodenham & Adams, 2016). Commonly, the approximation provided in Jensen and Solomon (1972) is used for obtaining the control limits in the context of process monitoring following Jackson & Mudholkar (1979).

2.2. CCA based monitoring

Consider a system with the mean-centred outputs Y, inputs $X \triangleq \{x_1, \dots, x_n \in \mathbb{R}^L, \dots, x_N\} \in \mathbb{R}^{L\times N}$, their respective sample covariance matrices $\tilde{\Sigma}_{yy} \succ 0$, $\tilde{\Sigma}_{xx} \succ 0$ and the cross covariance matrix $\tilde{\Sigma}_{yx}$, corresponding to the normal operation data. CCA extracts linearly independent latent variables from X and Y using the linear projection matrices $\zeta_x \in \mathbb{R}^{L\times J}$ and $\zeta_y \in \mathbb{R}^{P\times J}$, $J = \min(L, P)$, respectively, such that the extracted latent variables have maximum correlation. When $\tilde{\Sigma}_{xx}$ and $\tilde{\Sigma}_{yy}$ are invertible, it can be obtained as the following,

$$\zeta_y = \tilde{\Sigma}_{yy}^{-\frac{1}{2}} \beta_y, \;\; \zeta_x = \tilde{\Sigma}_{xx}^{-\frac{1}{2}} \beta_x \tag{7}$$

from the following decomposition,

$$\tilde{\Sigma}_{yy}^{-\frac{1}{2}} \tilde{\Sigma}_{yx} \tilde{\Sigma}_{xx}^{-\frac{1}{2}} = \beta_y \Gamma \beta_x^T \tag{8}$$

where $\mathbb{R}^{J\times J} \ni \Gamma \succeq 0$ is a diagonal matrix with the singular values, and the matrices $\beta_y \in \mathbb{R}^{P\times J}$ and $\beta_x \in \mathbb{R}^{L\times J}$ contain orthonormal eigenvectors spanning the basis of the row and the column spaces of the matrix on the left hand side of Eq. (8), respectively. Due to the