Journal of Process Control 40 (2016) 62–77

Contents lists available at ScienceDirect

Journal of Process Control

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

Bayesian robust linear dynamic system approach for dynamic process monitoring

Jinlin Zhu, Zhiqiang Ge*, Zhihuan Song

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 4 January 2016
Accepted 25 January 2016
Available online 19 February 2016

Keywords:
Robust linear dynamic system
Variational Bayesian method
Kalman filter
Student's *t*-distribution
Fault detection

ABSTRACT

In this paper, a Bayesian robust linear dynamic system approach is proposed for process modeling. Traditional linear dynamic system (LDS) constructed with Kalman filter is designed by Gaussian assumption which can be easily violated in non-Gaussian modeling situations, especially those with outliers. To deal with this issue, the conventional Gaussian-based Kalman filter is modified with heavy tailed Student's *t*-distribution so as to deal with the non-Gaussian noise and modeling outliers. Then, a variational Bayesian expectation maximization (VBEM) algorithm is developed for learning parameters of the robust linear dynamic system. For process monitoring, traditional monitoring scheme are discussed and the residual space monitoring mechanism has been improved. To explore the feasibility and effectiveness, the proposed method is applied for fault detection, with detailed comparative studies with several other methods through the Tennessee Eastman benchmark.

1. Introduction

Industrial processes have become more and more complex, with the development of modern manufacturing equipments and process control mechanisms [1,2]. In order to maintain the safe operation of process systems, process monitoring is particularly essential [3]. Among the various monitoring schemes, data driven multivariate statistical process monitoring (MSPM) models make little requirement for accurate kinematic equations and are easy to be established [4–6]. Consequently, MSPM models such as principal component analysis (PCA) are widely studied and applied over the past few decades [7].

PCA can detect faults effectively by constructing T^2 statistic using latent space information; meanwhile, *SPE* statistic is also constructed by the residual space so as to monitor the change of measurement noise. Although traditional PCA has been popular and effective for many industrial applications, one drawback is that the original deterministic model is lack of natural interpretation of the process uncertainties [8,9]. Thereby, probabilistic PCA (PPCA) has been proposed recently. On the one hand, the PPCA incorporates process uncertainties by employing Gaussian distributions on latent and observed modeling spaces; therefore, one can describe data variances in a more elegant manner. On the other hand, irregular process data such as missing items can also be well handled within expectation maximization learning mechanism [10]. With further improvement, PPCA has also been extended by mixture formulation so as to model those processes with multiple operating modes [11].

As static modeling methods, both PCA and PPCA assume data samples as independently collected from sensors with no time serial correlations. However, it is well known that most industrial processes evolve from past operation situations to potential future events [12,13]. Therefore, dynamic behavior should be one of the most important characteristics for industrial process data [14–19]. To consider the time serial related property, dynamic PCA (DPCA) has been developed by augmenting each measurement with a fixed length of several previous measurements and aligned to a stacking matrix [20]. After that, similar PCA projections and statistics are then constructed. Another commonly used dimensionality reduction based method is canonical variate analysis (CVA) [21]. CVA considers correlations by maximizing the related correlation index among variables, some studies show that compared with PCA and DPCA, CVA provides more desirable monitoring performance [22]. It should be noted that both DPCA and CVA are built with deterministic manner and no probabilistic

* Corresponding author. Tel.: +86 87951442.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

http://dx.doi.org/10.1016/j.jprocont.2016.01.010
0959-1524/

interpretations for noise uncertainties have been considered. As a probabilistic alternate, a data-based linear Gaussian state space model (LGSSM) has been developed [3]. LGSSM is a linear dynamic system (LDS) which constructs state space model with Gaussian states and Gaussian noises. Upon that, Kalman filter and EM iterations are introduced for estimating states and model parameters. Simulation results show that compared with traditional PCA and PPCA, LGSSM is more desirable for dynamic process modeling and monitoring [3]. However, one common issue for LDS method is that the Gaussian assumption can be violated by potentially sampling outliers [23]. Outliers can be hardly avoided and may cause model misspecification for Gaussian based methods [24]. The reason is that Gaussian models assume the distribution tails drop exponentially which is known as the three-Sigma principle [25]. However, outliers usually occur in or beyond the three-Sigma region. Even if outliers are embedded within such regions, outliers are not safe since the potential non-Gaussian variations may challenge the Gaussian assignation. In this condition, estimated significant parameters like mean and covariance can be skewed [26].

In this article, a Student's *t*-based assumption has been employed for LDS observation space so as to tolerate sampling outliers. Besides mean and scale matrix, the Student's *t*-distribution is defined with a tail adjust parameter called degree of freedom [27,28]. Thus, the Student's *t*-distribution provides reasonable tolerance for outliers without badly distorting the entire distribution [29]. In this sense, the derived method is called robust LDS. In addition to outlier modeling problem, another important issue is parameter learning. In order to make the closed form calculation such as maximum likelihood (ML) estimation computationally tractable, the original Student's *t*-distribution is usually represented as an infinite combination of Gaussian formulas with same mean and covariance parameters. However, several drawbacks can be found for ML methods [30]. First, the log-likelihoods in ML are not bounded which may result in singular covariance results [31]. Second, EM derived ML method can be easily stuck into local maxima [32]. As an alternative, variational Bayesian (VB) inference overcome these problems by treating model parameters as random quantities and taking integrations over these unknown quantities [33,34]. In other words, instead of finding a point estimation over parameters, VB based EM tries to derive some easy-to-handle approximate distributions over these parameters. Compared with ML based method, VB based EM algorithm can reach more appreciate solutions. Based on this, we propose a VB based EM algorithm for robust LDS modeling and the derived method is called variational Bayesian robust LDS (VBRLDS). In the VBE step, Kalman filter and smoother is employed for estimating expectation related terms, while in the VBM step, the variational inference is called so as to re-estimate those parameters with obtained expectation terms. Finally, in order to apply the VBRLDS for fault detection, traditional statistics are discussed and the control limit for SPE statistics has been modified so as to tolerate outliers. A two-component Gaussian mixture model based scheme has been proposed for designing control limit under the disturbance of outliers.

The remainder of this paper is organized as follows. In Section 2, robust LDS is first defined. Then, a variational Bayesian EM algorithm is proposed in Section 3, followed by the fault detection schemes and discussions. In Section 4, a comparative study is conducted. In the last section, conclusions are made.

2. Method

2.1. Robust linear dynamic system

The state space model for robust linear dynamic system can be defined as [3]

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \tag{2}$$

where $\mathbf{y}_k \in R^D$ is the D dimensional observation vector at time $k(k=2, 3, ..., N)$, $\mathbf{x}_k \in R^d$ is the d dimensional latent state vector, $\mathbf{A} \in R^{d\times d}$ denotes the state transition matrix, $\mathbf{C} \in R^{D\times d}$ is measurement matrix, $\mathbf{w}_k \in R^d$ is assumed as Gaussian distributed process noise, $\mathbf{v}_k \in R^D$ is assumed as Student's *t*-distributed non-Gaussian measurement noise with heavy tail. We have $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q})$ and $\mathbf{v}_k \sim t(0, \Sigma, \nu)$. Here $N(\cdot)$ and $t(\cdot)$ denote Gaussian and Student's *t* distributions, respectively, $\mathbf{Q}$ is diagonal state covariance matrix, Σ is a diagonal observation scale matrix and ν is degree of freedom. Hence, state transition and observation probabilities can be given as [3]:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathbf{N}(\mathbf{x}_k|\mathbf{A}\mathbf{x}_{k-1}, \mathbf{Q}) \tag{3}$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = t(\mathbf{y}_k|\mathbf{C}\mathbf{x}_k, \Sigma, \nu) \tag{4}$$

Please notice that the state space in Eq. (3) is still modeled by Gaussian distribution since we assume that outliers should be modeled and explained within measurement space which is a heavy tailed Student's *t* distribution as defined in Eq. (4). A D-dimensional Student's *t*-density with mean $\mathbf{x}_k$, scale matrix Σ and degree of freedom ν can be defined as [26]:

$$t(\mathbf{y}_k|\mathbf{C}\mathbf{x}_k, \Sigma, \nu) = \frac{\Gamma\left(\frac{D+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{D}{2}}}\left|\Sigma^{-1}\right|^{\frac{1}{2}}\left[1+\frac{1}{\nu}(\mathbf{y}_k-\mathbf{C}\mathbf{x}_k)^T\Sigma^{-1}(\mathbf{y}_k-\mathbf{C}\mathbf{x}_k)\right]^{-\frac{D+\nu}{2}} \tag{5}$$

where $\Gamma\left(\frac{\nu}{2}\right) = \int_0^\infty z^{\frac{\nu}{2}-1}e^{-z}dz$ is the gamma function and the distribution tail is regulated by ν, a small ν makes a heavier tail and the Student's *t*-distribution deteriorates to the Gaussian fashion as $\nu \to \infty$. As mentioned above, the original Student's *t* formulation may result in computationally intractable solutions. Fortunately, the above Eq. (5) can be further analyzed with Gaussian mixtures with Gamma prior as [26]

$$t(\mathbf{y}_k|\mathbf{C}\mathbf{x}_k, \Sigma, \nu) = \int_0^\infty N(\mathbf{y}_k|\mathbf{C}\mathbf{x}_k, u_k^{-1}\Sigma)Ga\left(u_k|\frac{\nu}{2}, \frac{\nu}{2}\right)du_k \tag{6}$$

where $Ga(\cdot)$ is the Gamma distribution. One can infer from Eq. (6) that a latent prior variable u_k is used to control the shape of distribution by regulating the variance term.

Notice that if one defines the initial distribution of latent state variable as $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{P}_1)$, the entire parameter set need to be estimated can be given as $\Theta = \left\{\mathbf{A}, \mathbf{C}, Q, \Sigma, \nu, \boldsymbol{\mu}_1, \mathbf{P}_1\right\}$. As a further illustration, a graphical representation of the RLDS is given in Fig. 1. One can see