Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-21 (2015) 619–626

Sparse Contribution Plot for Fault Diagnosis of Multimodal Chemical Processes ⋆

Chao Ning, Maoyin Chen, Donghua Zhou

Department of Automation, Tsinghua University, Beijing 100084, P.R.China. (e-mail: ningc12@mails.tsinghua.edu.cn, mychen@mail.tsinghua.edu.cn, zdh@mail.tsinghua.edu.cn).

Abstract: Chemical processes usually work under different operating modes to meet the market demand and to achieve higher profits. This necessitates investigating related algorithms and alarm systems for multimodal chemical processes. Although some research effort has been made to monitor multimodal processes, little attention was paid to fault diagnosis issue when addressing multiple modes. In this paper, we present both a label consistent dictionary learning (LCDL) based multimode process monitoring approach and sparse contribution plot (SpCP) for fault diagnosis. Firstly, a discriminative and reconstructive dictionary is obtained from normal historical process data via label consistent K-SVD algorithm. In addition, we augment the learned dictionary to get another dictionary, which consists of two blocks, one for multiple normal operating modes and another for faults. Then, during online monitoring period, a new sample is coded sparsely using the aforementioned augmented dictionary. After that, its dictionary reconstruction residual (DRR) is calculated for fault detection purpose. At last, a novel sparse contribution plot is proposed to figure out the root cause of the detected fault. The SpCP is better able to highlight the real cause with no ambiguity in that only a small fraction of variables' sparse contributions are nonzeros. The effectiveness of the proposed methodology is demonstrated by both a numerical simulation and a continuous stirred tank heater (CSTH) process.

Keywords: multimode; process monitoring; fault diagnosis; sparse representation; dictionary learning; sparse contribution plot.

1. INTRODUCTION

In big data era, more and more chemical factories are beginning to make full use of large volume of online measurements to monitor health status of their plants. The process monitoring techniques have been widely adopted to ensure a safe operation and to prevent large economic damage. Multivariate statistical process monitoring (MSPM) techniques, including typical methods like principal component analysis (PCA) and partial least squares (PLS), have been reported in many research literature and achieved success in real applications(MacGregor and Kourti [1995]; Qin [2003]; Wise and Gallagher [1996]; Tong and Crowe [1995]). Due to the complexity of chemical processes, some improvements on traditional MSPM techniques have been made (Cheng et al. [2010]; Ge and Song [2007]; Zhou et al. [2010]). However, determining the occurrence of anomaly is just a good start, and it is not the end of the story. To be more specific, operators need to figure out the reason for the detected fault, which is the primary objective of fault diagnosis or fault isolation, to bring the process into normal status. In order to tackle this problem, contribution plots were proposed in the literature (Miller et al. [1998]). One of the big advantage of contribution plots is that they did not presume any prior knowledge of faults. Reconstruction based contribution (RBC) was proposed and achieved better results than traditional contribution plots (Alcala and Qin [2009]). The main limitation of RBC is that it presumed the fault direction or fault subspace, which does not hold true in many situations. Other works employed similarity factors to diagnose faults (Raich and Cinar [1996]; Yoon and MacGregor [2001]; Kano et al. [2002]; Singhal and Seborg [2002]). While traditional fault detection and diagnosis methods have been well investigated, it should be noted that they did not consider multiple operating modes, which is a frequently encountered phenomenon (Hwang and Han [1999]).

In order to monitor multimode processes, researchers have proposed a lot of methods. One class of such methods, dubbed "model library based approach" (Ge et al. [2013]), models each operating mode separately. For example, the proposed Mix-PCA, multiple PCA and multiple PLS (Chen and Liu [1999]; Zhao et al. [2004]; Zhao et al. [2006]) belong to such category. More recently, hidden Markov model based statistics pattern analysis (HMM-SPA) was proposed for multimode process monitoring using an index-switching scheme (Ning et al. [2014]). There exists another class of methods, which employs combination strategy within Bayesian framework. For instance, A finite Gaussian mixture model (FGMM) and a Byesian inference based Gaussian mixture contribution (BIGMC) approach were presented (Yu and Qin [2008]; Yu [2013]).

⋆ This work was partially supported by the NSFC under grants 61210012, 61290324, 61490700, 61473164

10.1016/j.ifacol.2015.09.595

A mixture Bayesian regularization method of PPCA was developed in the literature (Ge and Song [2010]). Although there is a bulk of research literature addressing fault detection for multimode processes, little attention was paid to figuring out the reason for the detected fault.

Recently, sparse representation has gained tremendous attention due to its inspiring success in various fields, such as computer vision, image processing and pattern recognition (Elad and Aharon [2006]; Elad et al. [2010]; Wright et al. [2010]). In these studies, a collection of basis, often redundant, is used to construct a dictionary. And the high-dimensional data is assumed to be a compact representation of a small fraction of the basis but with high fidelity. Wright et al. proposed to use the whole training data set as a dictionary and put forward sparse representation classification (SRC) to perform robust face recognition (Wright et al. [2009]). However, in some real applications, the training data set tends to be large, which makes the computation expensive. Thus, some literature on sparse representation makes an attempt to learn a dictionary with an appropriate size from the training data instead of using the training data per se (Aharon et al. [2006]; Wang and Kong [2014]). Moreover, some research effort has been made to make the dictionary discriminative by taking advantage of the label information, leading to label consistent K-SVD algorithm (Jiang et al. [2013]). There emerges a surge of interest in sparse representation among fault detection researchers. An updated dictionary learning was proposed for oil pipeline leakage detection (Yan et al. [2011]). Indeed, the authors trained two dictionaries, i.e. 'negative' dictionary and 'positive' dictionary, from normal training data and faulty data respectively. And then they used coding residual to perform classification. However, this method is impractical in some industrial cases where the number of faulty data is far from sufficient to train the 'positive' dictionary. The SRC was employed to deal with multi-class classification for normal data and multiple types of faults, thus carrying out fault detection and diagnosis (FDD)(Wu et al. [2012]). The largest limitation of this method is that it assumed fault types are predefined and new types of faults will never occur. Ren and Lv proposed a fault detection algorithm via sparse representation (FD-SR) in semiconductor manufacturing (Ren and Lv [2014]). While sparse representation has achieved inspiring success in fault detection, in FDD research area, we are also interested in fault diagnosis, i.e. isolating the faulty variable.

The main contributions of this paper are listed as follows. First, in the proposed method, we do not presume the distribution of the process data within each operating mode. The robustness to the distribution is partly ascribed to the wise selection of the dictionary atoms (Ren and Lv [2014]). And the selection often depends on dictionary learning algorithms. Second, taking inspiration from the work (Wright et al. [2009]), a novel dictionary reconstruction residual (DRR) is proposed for multimode process monitoring. Although the idea behind DRR is similar to that in the literature (Ren and Lv [2014]), which assumed the normal process data can be represented sparsely with respect to the dictionary with small coding error whereas the faulty data can not be represented sparsely with high accuracy, the dictionary they used was not augmented and they did not provide fault diagnosis technique using their learned dictionary. Third, a novel sparse contribution plot is proposed in this paper. We inherit the merits of sparse representation, and develop a fault diagnosis method for multimode processes. Unlike in the literature (Wu et al. [2012]), we do not presume the knowledge about the types of faults or make the assumption that no novel fault will occur. To our best knowledge, there exists no literature dealing with fault diagnosis for multimode processes within the sparse representation framework.

The remainder of this paper is organized as follows. In Section 2, label consistent dictionary learning based fault detection method and sparse contribution plots for fault diagnosis are proposed within sparse representation framework. Section 3 provides a numerical simulation and a continuous stirred tank heater (CSTH) process, both of which are used to demonstrate the effectiveness and superiority of the proposed method. Section 4 draws up our conclusion.

2. FAULT DETECTION AND DIAGNOSIS FOR MULTIMODE PROCESSES VIA SPARSE REPRESENTATION

2.1 Offline Modeling: LCDL algorithm

In this section, we want to present the offline modeling procedure, in which a discriminative and reconstructive dictionary is learned from historical process datasets. The process operates at multiple modes, and samples generated from different modes have different labels. Therefore, a discriminative dictionary is appropriate to characterize the nature of multimodal process data. Besides, reconstruction errors are often utilized to detect anomaly, thus making it necessary for the learned dictionary to have a strong capability of reconstruction. Based on the above two observations, a label consistent dictionary is more suitable for fault detection in multimodal chemical processes.

Suppose there are C operational modes and $X = [X_1, \ldots, X_i, \ldots, X_C] \in \mathbb{R}^{m \times N}$ denotes the normal process dataset, where $X_i \in \mathbb{R}^{m \times N_i}$ is the collected data set from the i^{th} mode with m sensors and N_i samples. And the total number of normal samples is $N = \sum_{i=1}^{C} N_i$.

Each column of the overall data set X has a label, which is the operating mode this column belongs to. Define an objective function for dictionary learning as follows (Jiang et al. [2013]):

$$
\begin{aligned}
< D, U, W, A > = argmin \left\| X - DA \right\|_F^2 + \lambda \left\| Q - UA \right\|_F^2 \\
+ \beta \left\| L - WA \right\|_F^2 \\
s.t. \quad \forall i \ \left\| \alpha_i \right\|_0 \leq n_0
\end{aligned}
$$

where $D = [d_1, d_2, \cdots, d_K] \in \mathbb{R}^{m \times K}$ is the learned dictionary, $A = [\alpha_1, \alpha_2, \cdots, \alpha_N] \in \mathbb{R}^{K \times N}$ are the sparse codes of the training data X, $L = [l_1, l_2, \cdots, l_N] \in \mathbb{R}^{C \times N}$ are the labels of the training data X. If the training sample x_i is generated from the j^{th} operating mode, $l_i = [0, 0, \cdots, 1, \cdots, 0, 0]$ where only the j^{th} entry is 1. $Q = [q_1, q_2, \cdots, q_N] \in \mathbb{R}^{K \times N}$ are discriminative sparse codes of the training data. If the i^{th} training sample x_i