Chemometrics and Intelligent Laboratory Systems 157 (2016) 111–119

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Fault detection with improved principal component pursuit method

Yijun Pan, Chunjie Yang *, Ruqiao An, Youxian Sun

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

ARTICLE INFO

Article history:
Received 27 April 2016
Received in revised form 22 June 2016
Accepted 11 July 2016
Available online 13 July 2016

Keywords:
Fault detection
Blast furnace
Principal component pursuit
Low rank representation
TE process

ABSTRACT

In modern industries, principal component analysis (PCA) is one of the most popular data-driven methods. Since the directions of loading vectors can be severely affected by samples distribution, PCA is known for its sensitivity to outliers. In this paper, a robust principal component analysis method, improved principal component pursuit (IPCP), is proposed. Based on this IPCP method, a process model and an online monitoring statistic is developed. The traditional low rank representation (LRR) idea is introduced into PCP method. By applying this IPCP method, a low-rank coefficient matrix is constructed, and it represents explicit relationships between the variables as well as contains other useful information of the processes. Moreover, the obtained coefficient matrix is potentially useful to derive a powerful fault detection statistic. In order to test and evaluate the effectiveness of coefficient matrix constructed by IPCP and the power of proposed monitoring statistic, all algorithms are first tested in numerical simulation. Then they are illustrated in the TE process, which simulates the practical field condition. Finally the proposed methods are implemented in a blast furnace process.

1. Introduction

Recently, the control systems for industrial processes tend to be both large-scale and complex, and traditional fault detection methods designed for small-scale processes might not be effective. Therefore, it is important to develop novel data-driven fault detection methods suitable for the large-scale systems. The practical large-scale systems always work under high temperature, high pressure and high dust degree environments. Many anomalous events such as measurement errors, failures in data transmission and sudden changes in system behavior may lead to the presence of outliers [1]. Therefore, robust fault detection methods are necessary in industrial processes. Generally speaking, fault detection methods can be classified into three categories: model-based methods, field knowledge-based methods and data-based methods [2]. The most traditional one is model-based method, which is used for fault detection in aerospace, engine, and power systems. The model-based methods are based on exact process models, and they tend to give more accurate results than other two approaches as long as the system models are reliable. However, as modern industrial processes become more and more complicated, the characterization of first-principle models also becomes much more difficult, costly, and sometimes it is even impossible to build such models. The field knowledge-based methods are based on the expert experiences, so the monitoring results are more intuitive. However, these methods always spend much time and are difficult to operate requiring the long-term accumulation of expert knowledge and experiences. Compared to the other two approaches, by data-driven methods, the exact models of processes do not require to be built. Since distributed control systems are widely utilized in the modern industrial processes, a large amount of data has been collected, which contains abundant information about the industrial processes and thus is fairly useful for prompt fault detection [3–5].

One of the most widely used data-driven methods for instant faults detection is principal component analysis (PCA). However, it is well known that PCA is sensitive to outliers [6]. Recently, a new method of robust principal component analysis (RPCA) based on optimization, principal component pursuit (PCP), has been introduced by Candes et al. in 2009 [7]. Since then, PCP has attracted wide research interests due to its statistical and computational efficiency. For example, Bouwmans et al. reviewed the recent developments in the field of RPCA solved via PCP for the application of background/foreground separation [8]. Huan et al. proposed a novel PCP method based on sparse Bayesian learning principles and Markov random fields for image and video processing [9]. Bouwmans et al. also presented an overview of robust principal component analysis application, such as latent variable model selection, image processing, video processing and 3D computer vision [10]. Rodriguez et al. proposed a novel fully incremental PCP algorithm for video background modeling. This novel method has an extremely low memory footprint, and a computational complexity that allows real-time processing [11]. However, the method has been barely studied in the application of fault detection except for three papers: First,

* Corresponding author.
E-mail address: cjyang@iipc.zju.edu.cn (C. Yang).

http://dx.doi.org/10.1016/j.chemolab.2016.07.003
0169-7439/

Isom et al. indicated that PCP is more favorable than PCA in multiple aspects including robustness [12]. In addition, the objectives of model building, fault detection, fault isolation, and process reconstruction can be simultaneously accomplished by observing corresponding residual. This is the first time that the PCP-based method is used for process monitoring; Second, Cheng et al. made a further discussion on how to apply the PCP technique to process monitoring. A generalized median filter is proposed for analyzing each residual variable, which contains both outliers and faults [13]. Third, Yan et al. developed a robust process modeling and monitoring method based on stable PCP, and the $T^2$and SPE statistics of PCA are used for process monitoring [14].

There are two ways to deal with outliers based on PCA. The first one is to remove the outlying objects observed on the score plots and to repeat the PCA analysis again. Another way is more efficient to apply a robust, i.e. not sensitive to outliers, variant of PCA. To have a robust PCA, the simplest way to handle outliers is replacement of classical covariance or correlation matrix by a robust estimator [15]. Minimum covariance determinant (MCD) [16] and its fast version [17], S-estimators [18] and Minimum Volume Ellipsoid (MVE) are some of the most well-known estimators. Compared to other robust methods mentioned above, PCP leads to exact recovery of the underlying low-rank structure from a constant proportion of entry-wise corruptions with theoretical guarantees under weak conditions. Moreover, PCP can be implemented by applying fast numerical optimization algorithms such as alternating direction method of multipliers (ADMM) [7,19–21]. Therefore, it is very promising to apply PCP in modern fault detection for practical industrial processes.

In previous work, observing the residual in sparse matrix and the statistics of PCA are used for fault detection based on PCP method. However, the explicit statistics like T^2 and SPE in PCA should be developed in PCP method. Besides, although PCA and PCP are both dimensionality reduction method, PCP doesn't change matrix dimension. Therefore, the statistics of PCA may not be suitable to apply in PCP method directly. In this paper, a process modeling and an online statistic based on PCP are developed. The low rank representation (LRR) idea is introduced into PCP method for building a powerful statistic to fault detection. This method combines LRR idea with PCP idea to construct a new method, named improved PCP (IPCP). A stable PCP method applied for noises and outliers in processes is proposed by Zhou et al. [21]. Since the application of stable PCP for fault detection is similar to regular PCP and the magnitudes of noises are much smaller than outliers, the influence of noises is ignored to build a model in this paper for demonstrating conveniently. In industrial processes, some variables are used for describing one state to enhance reliability, so the sample variables are relevant [22]. By applying this IPCP method, a low-rank coefficient matrix is constructed, and it represents explicit relationships between the variables as well as contains other useful information of the processes. Therefore, the obtained coefficient matrix is potentially useful to derive a powerful fault detection statistic. The low-rank coefficient matrix is constructed by training matrix, and each sample of testing matrix is projected into this low-rank coefficient matrix to obtain a monitoring statistic. In order to test and evaluate the effectiveness of coefficient matrix constructed by IPCP and the power of proposed monitoring statistic, all algorithms are first tested in numerical simulation. Then they are illustrated in TE process, which simulates the practical field condition. Finally the proposed methods are implemented in practical circumstances. The simulations demonstrated the proposed methods are tested on a standard desktop computer with 1.86 GHz CPU and 2 GB of memory.

In Section 2, brief descriptions of PCP and LRR are given. The proposed monitoring model and statistic based on IPCP for process monitoring will be presented in Section 3. Section 4 provides the simulation results in numerical simulation, TE process and real blast furnace process respectively. Section 5 gives the conclusions.

2. Theory

2.1. Principal component pursuit (PCP)

Suppose we are given a data matrix $X \in R^{n \times m}$ with n observations and m measurement variables, and know that X may be decomposed by PCP as formula (1),

$$X = A + E \tag{1}$$

where A has low rank and E is sparse. $E_{ij} \neq 0$ means the ith observation of jth variable consisting of outliers. The theory of PCP is to solve the optimization problem (2),

$$\begin{aligned} &\min \|A\|_* + \lambda \|E\|_1 \\ &s.t.\ \ X = A + E \end{aligned} \tag{2}$$

where $\|A\|_*$ denote the nuclear norm of the matrix A, and it is equal to the sum of singular values. $\|E\|_1$ denote the l_1 norm of the matrix E, and it is equal to the sum of the absolute value of all elements [7,23]. One might have expected that λ is a parameter used for balancing the two terms in $\|A\|_* + \lambda\|E\|_1$ appropriately. However, the $\lambda = \frac{1}{\sqrt{\max(n,m)}}$ is a correct choice no matter what A and E are. Therefore, a rather remarkable fact is that there is no tuning parameter in PCP method. Under the assumption of the theorem, choosing $\lambda = \frac{1}{\sqrt{\max(n,m)}}$ always returns the correct answer. The mathematical analysis can be found in reference [7].

2.2. Low rank representation (LRR)

Consider a set of data vectors $X = [x_1, x_2, \ldots, x_n]$ (each column is a sample) in R^D, each of which can be represented by the linear combination of the basis in a “dictionary” $A = [a_1, a_2, \ldots, a_n]$:

$$X = AZ \tag{3}$$

where $Z = [z_1, z_2, \ldots, z_n]$ is the coefficient matrix with each z_i being the representation of x_i. As it was mentioned, low rankness may be a more appropriate criterion capturing the global structures of the data X. So they look for a representation Z by solving the problem.

$$\begin{aligned} &\min\ \ rank(Z) \\ &s.t\ \ X = XZ \end{aligned} \tag{4}$$

In order to express the essence relationships of variables and segment the data into their respective subspaces, an affinity matrix is computed that encodes the pairwise affinities between data vectors. So the data X itself is used as the dictionary, i.e., problem (4) can be rewrote as formula (5) [24].

$$\begin{aligned} &\min \quad rank(Z) \\ &s.t. \quad\ X = XZ \end{aligned} \tag{5}$$

3. Fault detection bases on improved principal component pursuit method

3.1. Improved principal component pursuit

sAs mentioned above, the coefficient matrix Z is the “lowest-rank representation” of data X with respect to a matrix X itself. So this matrix Z represents explicit relationships between the variables of matrix X,