Chemometrics and Intelligent Laboratory Systems 149 (2015) 1–11

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Statistical process monitoring with integration of data projection and one-class classification

Yiqi Liu [a], Yongping Pan [b], Qilin Wang [c], Daoping Huang [a,*]

[a] School of Automation Science & Engineering, South China University of Technology, Wushang Road, Guang Zhou 510640, China
[b] Department of Biomedical Engineering, National University of Singapore, Singapore Medical Drive, 117575, Singapore
[c] Advanced Water Management Centre, The University of QLD, St. Lucia, Brisbane, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form 12 August 2015
Accepted 13 August 2015
Available online 14 September 2015

Keywords:
One-class classification
PCA
VBPCA
Process monitoring

ABSTRACT

One-class classification (OCC) has attracted a great deal of attentions from various disciplines. Few attempts are made to extend the scope of such application for process monitoring. In the present work, the Principal Component Analysis (PCA) and Variational Bayesian Principal Component Analysis (VBPCA) approach provides a powerful tool to project original data into lower data set as well as spreading different types of faults with different directions. This, along with multiple types of one-class classifiers (density-based, boundary-based, reconstruction-based and combination-based) that are able to isolate abnormal data from normal one, supported the design of process monitoring. These methodologies have been validated by process data collected from a Wastewater Treatment Plant (WWTP). The results showed that the proposed methodology is capable of detecting sensor faults and process faults with good accuracy under different scenarios.

1. Introduction

Due to the increasing safety requirement in the chemical processes, there is a strong need to monitor the operation process aiming to detect early and identify the root cause of a fault (fault diagnosis) or the abnormality that might negatively affect the process [1,2].

In the early stage, fault detection was performed by using univariate statistical techniques like simple thresholding [3]. However, since the industrial products are more heavily instrumented and complex increasingly, the data obtained is seldom univariate, but rather presents multivariate features. The resulted huge amount of correlated data requires proper techniques to extract useful information. One of the plausible ways is to project the data into a lower-dimensional space that accurately characterizes the state of the analyzed system. In this respect, Multivariate Statistical Process Control (MSPC) methods, such as principal component analysis [4,5], partial least-squares (PLS) [6], Gaussian process latent variable models [7] and independent component analysis (ICA) [8], are always used to make data more comprehensible by extracting relevant information. Also, they are capable of providing graphical tools easy to apply and interpret by process operators. However, traditional multivariate quality control charts, such as Hotelling's T^2 chart, assume that quality characteristics follow a multivariate normal distribution [9]. In many industrial applications, the distribution concerning the sampled data is not known, implying the need to construct a robust control chart suitable for the real applications. Even though proper control chart can be constructed, any pure data-based (or statistical-based) fault diagnosis methods will have difficulty in diagnosing on-line faulty data with much different faulty data.

As the limitations of current MSPC techniques, a more widely used approach applied to address unknown distributions consists of adjusting the control limits of conventional control charts by estimating the empirical percentiles of the monitored statistic. If the number of samples for training is small, the bootstrap re-sampling technique can be used [10]. Alternatively, a different paradigm consists of using classification methods that are designed to adapt data mining and machine learning technique to MSPC [11]. Up to date, both of non-linear classification, like artificial neural networks (ANN) [12] and Support Vector Machine (SVM) [13], and linear classification, like Fisher discriminate analysis (FDA), have been available for fault diagnosis. Most conventional classifiers assume more or less equally balanced data classes. However, for most of cases, they do not work well when any class is severely under-sampled or is completely absent, which is very common for a faulty class. One-class classification provide an alternative able to estimate a multivariate region that envelops the majority of training data. Then, the contour of this region plays the role of the control limit for further validation, and any observation that belongs to the minority and falls outside of this region is signaled. Therefore, these methods can be implemented without the need for prior knowledge about the data distributions. Several studies have been devoted such work recently with the goal of implementing one-class-classification algorithms as an alternative to traditional control charts [14,15], even as an extension to prediction modeling validation [16]. Sun and Tsung proposed

* Corresponding author. Tel.: +86 13610008387; fax: +86 20 87114189.
E-mail addresses: aulyq@scut.edu.com (Y. Liu), yongppan@gmail.com (Y. Pan), q.wang@awmc.uq.edu.au (Q. Wang), audhuang@scut.edu.com (D. Huang).

http://dx.doi.org/10.1016/j.chemolab.2015.08.012
0169-7439/

kernel-distance-based charts (K-charts) based on Support Vector Data Description (SVDD) algorithm, revealing that K charts performed better than T^2 charts [17]. Kumar et al. used one-class SVM to construct robust K charts through normalized monitoring statistics [18]. However, most of these control charts are limited to SVDD, i.e., boundary-based one-class-classification. Even though SVDD is able to perform well, the user should choose the kernel function, soft margin factor, and other corresponding parameters like Gaussian kernel distance, polynomial order, etc. The best selection of model parameters in SVM algorithm is not a trivial task and could lead to biased results. The aforementioned algorithms are all limited one type of OCC methods. To further gain more potentials for process monitoring, more OCC algorithms, such as density-based (Gaussian, mixture of Gaussians and Parzen density), reconstruction-based (PCA, k-means) [19] and combination-type OCC (mean, product and minimize) [20], are explored necessarily.

Another important issue that has largely ignored by fault diagnosis researchers is how to handle fault diagnosis in the presence of missing values. Generally, mean imputation is used to substitute missing data. To further improve the imputation accuracy, Khan et al. address this issue of handling missing data in the target using Bayesian imputation and EM [21]. However, the aforementioned work is premised on the assumption that the dimensions of missing data are not more than two; in other words, only one sensor failure is permitted every time step. To recover missing data in the multivariate data sets, Arteaga and Ferrer gave a detailed analysis, showing that three different techniques, two projection based techniques and a third regression based one, could be used to handle this problem [4]. Even though the regression based technique prevailed in the past few years [22], projection based techniques still have potentials to be explored. Therefore, we extended VBPCA for data projection and missing data reconstruction. It can cope with the over-fitting problem by penalizing parameter values which correspond to more complex explanations of the data. Furthermore, using Bayesian learning to identify PCA parameters, missing data can be reconstructed smoothly [23].

This paper explores the possibility of integration of data projection methodologies with multiple types of one-class classifiers. We first proposed two efficient ways (PCA and VBPCA) to deal with high dimension issues in the data set. The presented methodologies can be shown to be useful not only for dimension reduction, but also for fault direction isolation. Also, due to the positive performance of Variational Learning in the PCA identification, original data is able to be reduced into lower dimension even they were contaminated by missing values. The extension of multiple one-class-classification, including density-based (Gaussian, mixture of Gaussians, Parzen density), reconstruction-based (PCA, k-means), boundary-based (NNDD, SVDD) and combination-based methods, were used to establish the control limits necessary to improve the existing fault diagnosis methods. The resulting fault diagnosis methods with the mixture of projection methods and one-class-classification were able to identify multiple faults, rather than a single fault as conventional MSPC.

The remaining sections of this paper are organized as follows. Section 2 gives some basic theories on PCA and VBPCA. Section 3 discusses the implementation of an integration of projection methods with multiple one-class classifications for process monitoring. Section 4 presents the performance of the proposed methods through a wastewater treatment process. Some results are discussed in Section 5. Finally, the work is concluded in Section 6.

2. Preliminaries

2.1. Dimension reduction methods

Modern industrial processes often present huge amounts of process data due to the large number of frequently measured variables. One of the most widely used methods to deal with this problem in industries is PCA, which is able to compress high-dimensional data into a lower-dimensional space, thus, making data more comprehensible by extracting essential information. PCA uncover combinations of the original variables with orthogonal transformation to a set of values of linearly uncorrelated variables called principal components (PC). Due to its sensitivity to the relative scaling of the original variables, PCA is always used for fault detection and diagnosis in industries [24,25].

Principal component analysis (PCA) is a classical data analysis technique and widely used for fault diagnosis. However, over-fitting and bad locally optimal solutions could be expected in the presence of missing data. The VBPCA provides a good alternative for handling missing values. The basic idea behind VBPCA is to recast the problem of computing posterior probabilities-which is inherently a high-dimensional integration problem-as an optimization problem by introducing a class of approximating distributions, then optimizing some criterion to find the distribution within this class that best matches the posterior. For VBPCA, it is to find the most probable parameter set $\theta = \{W, v_x\}$ in the model structure:

$$x = Wt + e \tag{1}$$

where $x \in R^{d \times n}$ is the raw data $W \in R^{d \times d}$ denotes loading matrix. d is dimensionality of the data vectors, n is number of data vectors. Both the principal components t and the noise e are assumed normally distributed:

$$p(t) = N(t : 0, I) \tag{2}$$

$$p(e) = N(e : 0, v_x I) \tag{3}$$

where v_x is noise variance and I is the unit matrix. PCA is actually a special case of VBPCA model with v_x being restricted to 1. The parameter set θ can be estimated by the Variational Bayesian learning as illustrated in the following section. The core in VBPCA is the evaluation of the probability densities of all variables in the model. $p(t)$, $p(e)$, $p(x)$ and also posteriors $p(t|x)$, $p(e|x)$. Since $p(t_j) = N(t_j : 0, I)$ and $p(e_j) = N(e_j : 0, v_x I)$ are already assumed in VBPCA model, the remaining variables are $p(x)$, $p(t|x)$ and $p(e|x)$, it is obvious that $p(x)$ subject to normal distribution:

$$N\left(x : 0, WW^T + v_x I\right) \tag{4}$$

where $p(t|x) = N(t|x : \mu^{t|x}, \sigma^{t|x})$, posterior mean and posterior covariance are computed as follow $\mu^{t|x} = (W^T W + v_x I)^{-1} W^T x$, $\Sigma^{t|x} = v_x (W^T W + v_x I)^{-1}$ [26]. Let $M = (W^T W + v_x I)^{-1}$, we can reformulate it as

$$N\left(t|x : MW^T x, v_x M\right). \tag{5}$$

According to the conditional Gaussian property, $(e|x) = N(e|x : \mu^{e|x}, \sigma^{e|x})$ where $\mu^{e|x} = (I - WMW^T)x$ and $\sigma^{e|x} = v_x WMW^T$, thus,

$$N\left(e|x : \left(I - WMW^T\right)x, v_x WMW^T\right). \tag{6}$$

It is obvious that the observations can be decomposed to its systematic part and noise part for further fault diagnosis:

$$x = WE[t|x] + E[e|x] = \left(WMW^T\right)x + \left(I - WMW^T\right)x \tag{7}$$

where $x = E[x|x]$.

All the computation aforementioned are totally dependent on the adequate estimate of the model parameters, W and v_x. To solve this problem, Variational Bayesian (VB) learning is performed to calculate the cost function [27]. VB learning is sensitive to posterior probability mass rather than posterior probability density, thus making it more resistant against over-fitting compared to other estimation methods (Least squares, Point estimation) [19]. Typically, over-fitted solutions