Chemometrics and Intelligent Laboratory Systems 130 (2014) 20–28

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Statistical process monitoring based on a multi-manifold projection algorithm

Chudong Tong, Xuefeng Yan *

Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 6 April 2013
Received in revised form 28 August 2013
Accepted 15 September 2013
Available online 20 September 2013

Keywords:
Process monitoring
Manifold learning
Feature extraction
Principal component analysis

ABSTRACT

Considering that the global and local structures of process data would probably be changed in some abnormal states, a multi-manifold projection (MMP) algorithm for process monitoring and fault diagnosis is proposed under the graph embedded learning framework. To exploit the underlying geometrical structure that contains both global and local information of sampled data, the global graph and local graph are designed to characterize the global and local structures, respectively. A unified optimization framework, i.e. global graph maximum and local graph minimum, is then constructed to extract meaningful low-dimensional representations for high-dimensional process data. In the proposed MMP, the neighborhood embedding is used in both global and local graphs and the extracted features are faithful representations of the original data. The feasibility and validity of the MMP-based process monitoring scheme are investigated through two case studies: a simple simulation process and the Tennessee Eastman process. The experimental results demonstrate that the whole performance of MMP is better than those of some traditional preserving global or local or global and local feature methods.

1. Introduction

The increasing demands for production safety and quality stabilization of chemical plants continue to draw attention to research in process monitoring. Notable advancement of computing and information technology over the past decades has led to wide application of distributed control system, and thus a large number of data are stored for the chemical processes, which is the solid foundation of multivariate statistical process monitoring (MSPM). Principal component analysis (PCA) and partial least squares (PLS) are two major MSPM methods and have been actively investigated [1–3]. Both PCA and PLS project the highly correlated and noisy data onto a low-dimensional subspace that best characterizes the variances of the original data space. In other words, the global structure of dataset is preserved. Meanwhile, several extensions of PCA and PLS have been extensively reported in the literature, each of which is characterized by focusing on specific process aspects in order to get better monitoring performance. Dynamic PCA and PLS use a time lag shift method to include time-variant property in dynamic processes [4,5]. Lee et al. proposed a nonlinear process monitoring scheme based on kernel principal component analysis (KPCA) and achieved certain success [6]. Besides, adaptive techniques are developed recently to update the model consistently in order to monitor the process with changing conditions [7,8]. Recently, an important complementary statistical process monitoring scheme based on independent component analysis (ICA) has been proposed for non-Gaussian process monitoring and fault diagnosis [9,10]. ICA seeks to decompose the original dataset into linear combinations of statistically independent components and to deal with higher-order statistics. Ge and Song proposed a unified framework for MSPM based on independent component analysis–principal component analysis (ICA–PCA) [11]. ICA–PCA functions very well since it combines the advantages of both PCA and ICA. Generally, the essence of those methods is to conduct dimensionality reduction and find a reduced space where the feature of the original data can be faithfully represented. Therefore, different dimensionality reduction techniques identify different features of the original dataset, based on which the monitoring performance would also be affected differently.

It is well known that PCA can easily handle high-dimensional, noisy, and highly correlated data generated from chemical processes. However, it is worthwhile to stress that the PCA model only considers the global structure, and the detailed local structure feature is ignored [12–15]. In the absence of inner data structure analysis, only limited information can be extracted from process dataset. Recently, a new dimensionality reduction technique known as manifold learning has been gaining much attention for preserving local structure in pattern recognition area. Several algorithms focused on local structure preserving, such as locally linear embedding (LLE) [13], Laplacian eigenmaps (LE) [14], and locality preserving projections (LPP) [15], have been proposed to exploit the underlying geometrical manifold of dataset. Experiments have shown that these methods can find perceptually meaningful embedding for artificial and real-world datasets [16]. In

* Corresponding author at: East China University of Science and Technology, P.O. BOX 293, MeiLong Road NO. 130, Shanghai 200237, PR China. Tel./fax: +86 21 64251036.
E-mail address: xfyan@ecust.edu.cn (X. Yan).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.09.006

contrast to other manifold learning algorithms, LPP can possess a remarkable advantage that it can generate an explicit map, which is linear and can be easily obtained like PCA. More recently, some extensions of the LPP algorithm were also proposed for pattern recognition to overcome certain limitations of the original LPP and achieved better recognition performance [17–19]. The basic idea of LPP is to find an optimal projection such that the local neighborhood structure can be preserved in a low-dimensional space [15]. LPP inherits local structure preserving characteristic and has been successfully applied to fault detection of batch process and nonlinear process [20,21]. However, the dimension reduction performance of LPP could be degraded for losing some important information existing in global structure of given observations [22].

In general, either the global structure or the local structure of normal dataset is changed after an abnormal event occurs. Therefore, the global and local structures are crucial for process monitoring since the global structure defines the outer shape of the process dataset and the local structure presents inner organization. A natural idea is to combine the two phases together in order to improve the fault detection performance. Fortunately, Zhang et al. proposed a global–local structure analysis (GLSA) model for fault detection and it outperforms PCA and LPP-based monitoring methods [22]. In addition, Yu proposed a similar monitoring scheme based on local and global PCA (LGPCA) models [23] afterward. As a compromise, GLSA and LGPCA preserve both global and local structures through a simple combination of PCA and LPP. The main difference of these two methods locates in the way they integrate cost functions of PCA and LPP.

Motivated by the neighborhood embedding ability of manifold learning, a novel dimensionality reduction algorithm, called multi-manifold projection (MMP) is proposed in the current work. MMP aims to find a low-dimensional representation which can optimally preserve the global structure and the local structure, simultaneously. The basis of MMP is that the neighborhood information is embedded in both global and local graphs, which is partially ignored by GLSA and LGPCA. The feasibility and validity of MMP-based process monitoring scheme are illustrated through two case studies: a simple numerical simulation and the Tennessee Eastman process (TEP), and the experimental results demonstrate that the MMP-based method outperforms PCA-based and LPP-based monitoring methods and gives better monitoring performance than that of GLSA and LGPCA.

The remainder of this paper is organized as follows. Firstly, MMP algorithm derivation is presented in Section 2. Section 3 provides the algorithm analysis and the MMP-based process monitoring scheme. Two simulated processes are then used to illustrate the effectiveness of the proposed method in Section 4. Finally, conclusions and extensive discussions are given at the end of the article.

2. Algorithm formulation

Denote the dataset as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]^T$, $\mathbf{x}_i \in \mathrm{R}^m$. The problem of dimensionality reduction is to find a projection $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_d]$ which maps $\mathbf{X}$ to a low-dimensional sample set $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$, $\mathbf{y}_i \in \mathrm{R}^d$, $d \ll m$, such that the original data is faithfully represented.

2.1. PCA and LPP

The objective of PCA is to find a projection axis $\mathbf{p}$, such that the projected data variance is maximal, which is given as follows:

$$J_{PCA} = \max \sum_{i=1}^{n} \mathbf{p}^T (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \mathbf{p} \tag{1}$$

where $\overline{\mathbf{x}} = (1/n)\sum \mathbf{x}_i$. The low-dimensional sample points $\mathbf{y}_i = \mathbf{p}^T \mathbf{x}_i$ have the same directions of maximal variance with the original dataset $\mathbf{X}$. PCA considers only the outer shape of the given data but lacks the ability to extract the local representations. Without considering the local relationship between all pairs of data points, PCA destroys the intrinsic geometrical structure of the dataset.

In many real-world industrial processes, the sampled dataset is often with complicated distributions, and thus the local structure preserving is very important. Fortunately, LPP was proposed to tackle this problem, and the following objective function is used [15]:

$$J_{LPP} = \min \sum_{i,j=1}^{n} \mathbf{p}^T \left(\mathbf{x}_i - \mathbf{x}_j\right) W_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{p} \tag{2}$$

where $\mathbf{W}$ is a $n \times n$ an adjacency matrix with elements calculated by [13,15]:

$$W_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2 / c\right) & \text{if } x_i \in N\left(x_i; x_j\right) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $N(\mathbf{x}_i, \mathbf{x}_j)$ denotes that $\mathbf{x}_i$ is among k nearest neighbors of $\mathbf{x}_j$ or $\mathbf{x}_j$ is among k nearest neighbors of $\mathbf{x}_i$, and c is a parameter for adjusting W_{ij}. The value of W_{ij} represents the neighborhood relationship between $\mathbf{x}_i$ and $\mathbf{x}_j$. The projection axis $\mathbf{p}$ minimizing the objective function of LPP can preserve the local information of the dataset $\mathbf{X}$. However, without respect to the faraway data points that represent outer shape in the space $\mathbf{X}$, LPP may lose the variance information and the outer shape of the dataset may be destroyed.

2.2. Multi-manifold projection (MMP)

In order to overcome the shortcomings mentioned above, the neighborhood structure of the data is embedded in both local and global information. By taking the advantages of neighborhood embedding ability of manifold, a unified dual optimization function is constructed for the proposed MMP algorithm. To this end, two types of graphs in MMP: local graph and global graph are defined to simplify statements, respectively.

2.2.1. Local graph minimum

For local graph, an adjacency matrix $\mathbf{W}$ is first calculated. The locality preserving criterion is given as follows [15]:

$$\begin{aligned} J(\mathbf{p}) &= \min \sum_{i,j=1}^{n} \mathbf{p}^T \left(\mathbf{x}_i - \mathbf{x}_j\right) W_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{p} \\ &= \min \mathbf{p}^T \mathbf{X}^T (\mathbf{D} - \mathbf{W}) \mathbf{X} \mathbf{p} \\ &= \min \mathbf{p}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{p}^T = \min \mathbf{p}^T \mathbf{L}' \mathbf{p}^T \end{aligned} \tag{4}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is known as Laplacian matrix in manifold learning, $\mathbf{L}' = \mathbf{X}^T \mathbf{L} \mathbf{X}$ is defined as local graph matrix, $\mathbf{D}$ is a diagonal matrix with diagonal elements being the column (or row) sum of $\mathbf{W}$, i.e. $D_{ii} = \sum_j W_{ij}$, and D_{ii} represents a point's nearby density.

2.2.2. Global graph maximum

For global graph, it would be necessary to embed the neighborhood information to obtain an optimal outer shape manifold structure. Here, the local mean center of each sample $\mathbf{x}_i$ is considered, which can be more respective than the original mean center $\overline{\mathbf{x}}$. As mentioned previously, the diagonal elements in $\mathbf{D}$ reveal the nearby density of corresponding points. According to $\mathbf{W}$, the local mean vector of $\mathbf{x}_i$ is given by:

$$\overline{\mathbf{x}}_i = \frac{1}{n_i} \sum\nolimits_{\mathbf{x}_j \in N\left(\mathbf{x}_i, \mathbf{x}_j\right)} \mathbf{x}_j, i = 1, 2, \cdots, n \tag{5}$$