G Model
JJPC-2220; No. of Pages 7
ARTICLE IN PRESS

Journal of Process Control xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Process Control

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

Research Paper

Statistical process monitoring based on nonlocal and multiple neighborhoods preserving embedding model

Chudong Tong [a,*], Ting Lan [a], Xuhua Shi [a], Yuwei Chen [b]

[a] *Faculty of Electrical Engineering & Computer Science, Ningbo University, Ningbo 315211, PR China*
[b] *Key Laboratory of Rubber-Plastics, Ministry of Education/Shandong Provincial Key Laboratory of Rubber-Plastics, Qingdao University of Science & Technology, Qingdao 266042, PR China*

ARTICLE INFO

Article history:
Received 16 May 2017
Received in revised form 17 October 2017
Accepted 18 October 2017
Available online xxx

Keywords:
Neighborhood preserving embedding
Statistical process monitoring
Dimensionality reduction
Fault detection

ABSTRACT

A novel dimensionality reduction algorithm named nonlocal and multiple neighborhoods preserving embedding (NoMNPE) is proposed for modeling and monitoring industrial processes. The NoMNPE method implements dimensionality reduction by maximizing the variance scattered by nonlocal data points, while simultaneously preserving multiple neighborhoods relationships, which include time neighbors, distance neighbors, and angle neighbors for a given dataset. Therefore, three different manifold characteristics and one additional nonlocal relationship are taken into account in the NoMNPE model. The NoMNPE thus is expected to explore more intrinsic information in contrast to its counterparts, and could achieve enhanced monitoring performance as a result. The comparison studies on two industrial processes have also demonstrated the effectiveness and advantages of the proposed NoMNPE-based process monitoring approach.

1. Introduction

Modern industrial plants have been witnessing a wider application of computer-aided systems, and their related performance requirements thus motivate new approaches for efficiently and trustfully monitoring the operating condition for producing qualified products as well as ensuring plant stability. Among the diverse process monitoring methods, data-driven process monitoring, in particular, multivariate statistical process monitoring (MSPM) techniques have been intensively investigated [1–3]. For example, principal component analysis (PCA), independent component analysis, and partial least squares are mostly cited modeling algorithms. The design of a process monitoring system usually begins with a fault detection model that can well characterize the normal variation. Generally, a majority of MSPM methods implements dimensionality reduction to derive a reduced feature subspace that preserves the major important statistical information of the high-dimensional and collinear process data [4,5]. There are also some methods reveal normal process features without subspace extraction [6]. However, the efficiency of a fault detection model is still influenced by the properties of transformed components.

The purpose of PCA model is to explore the majority of data variance, which represents the global variability of the process data. From the geometrical viewpoint, the information of detailed local neighborhood structure is not taken into account in the formulation of PCA algorithm. However, the neighborhood information on the data manifold is also useful in designing a fault detection model since an abnormality in process operation can lead to a distortion in the underlying structure of sampled data. As a consequence, the PCA-based fault detection model cannot be always functional in detecting some specific faults given the loss of this important information. Fortunately, there are some dimensionality reduction algorithms focusing on data manifold available in the literature, such as locally linear embedding (LLE) [7], locality preserving projections (LPP) [8], neighborhood preserving embedding (NPE) [9] and so on. As a linear approximation to LLE, the NPE algorithm finds wider application in process monitoring through exploring the local geometrical structure of a given dataset [10,11]. Furthermore, the formulation of NPE algorithm makes it less sensitive to noise compared with the PCA algorithm [9]. However, the NPE-based fault detection methods assume that measurements are fairly independent over time. Without respect to the autocorrelation, applying NPE to dynamically correlated data would potentially have high missed detection rates.

To deal with the autocorrelated samples, one could refer to a lagged version of NPE to process multivariate variables with dynamic property, which is similar to the dynamic PCA (DPCA) method proposed by Ku et al. [12]. Although an augmented data

* Corresponding author.
E-mail addresses: tongchudong@nbu.edu.cn, cd-tong@hotmail.com (C. Tong).

https://doi.org/10.1016/j.jprocont.2017.10.009
0959-1524/

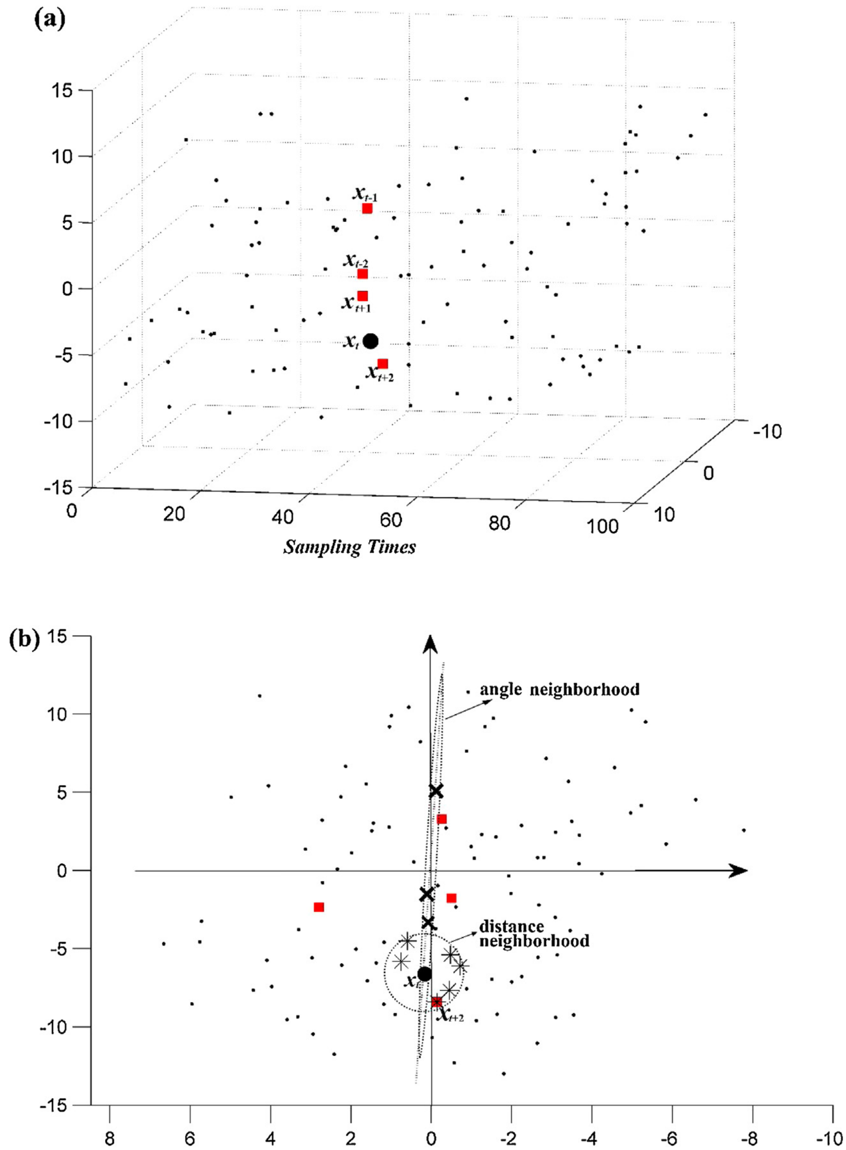

Fig. 1. Illustration of (a) time neighbors, (b) distance neighbors and angle neighbors.

matrix including lagged measurements seems to be the simplest way to model the data autocorrelation, the dynamic NPE might misinterpret the intrinsic geometrical proximity relations of the training data since it treats the lagged measurements to be a single data point. Recently, Li et al. [13] developed a dynamic latent-variable (DLV) model to improve the interpretation of dynamic processes and to enhance the dynamic process monitoring performance. The DLV can extract dynamic latent factors that are dynamically correlated and static latent factors that are statistically time-independent explicitly. Furthermore, Miao et al. [14] proposed a time neighborhood preserving embedding (TNPE) model by reconstructing each sample from its sampling time neighbors instead of distance neighbors. The consideration of time-series relationship of a data manifold provides an additional alternative for the dynamic process modeling. Although the TNPE could provide better monitoring performance than its counterparts, such as DPCA, NPE, and dynamic NPE, in monitoring dynamic processes, the loss of the consideration of distance neighborhood would still leave some specific faults remain undiscovered. Additionally, the achieved monitoring performance is sensitive to the pre-determined number of time neighbors, which will be shown in the case study section.

An additional shortcoming of TNPE is that the nonlocal relations of the training data have not been modelled. An examination of the existing literature, however, demonstrates that the simultaneous consideration of locality and nonlocality could result in enhanced monitoring performance [15–17]. Recognition of the issues mentioned above motivates us to propose a nonlocal and multiple neighborhoods preserving embedding (NoMNPE) model for data-driven process monitoring. The NoMNPE model implements dimensionality reduction by maximizing the nonlocal data points as well as preserving multiple neighborhoods relationships, which include time neighbors, distance neighbors, and angle neighbors. The Fig. 1 presents a two-dimensional illustration of the time neighbors, distance neighbors, and angle neighbors for a data point $\mathbf{x}_t$. The time neighbors of $\mathbf{x}_t$ are marked by red square points as shown in Fig. 1(a), the order of sampling times is used to determine the time neighbors. The Fig. 1(b) displays the distance neighbors and angle neighbors connected to $\mathbf{x}_t$, which are marked by asterisk points and cross points, respectively. In addition, some neighbors can have multiple attributes. For example, the marked point $\mathbf{x}_{t+2}$ as shown in Fig. 1(b) can be time neighbor and also distance neighbor of $\mathbf{x}_t$. The nonlocal neighbors are those data points having no