Chemometrics and Intelligent Laboratory Systems 157 (2016) 177–188

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Nonlocal and local structure preserving projection and its application to fault detection

Lijia Luo *, Shiyi Bao, Jianfeng Mao, Di Tang

Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Engineering Research Center of Process Equipment and Remanufacturing, Ministry of Education, Hangzhou 310014, China

ARTICLE INFO

Article history:
Received 5 May 2016
Received in revised form 28 July 2016
Accepted 31 July 2016
Available online 02 August 2016

Keywords:
Process monitoring
Fault detection
NLLSPP
Dimensionality reduction
Nonlocal and local data structures

ABSTRACT

A novel dimensionality reduction method named nonlocal and local structure preserving projection (NLLSPP) is proposed and used for process monitoring. NLLSPP can simultaneously preserve the nonlocal structure (i.e., data variance) and the local structure (i.e., neighborhood relationships between data points) of the data set. According to nonadjacent or neighboring relationships of different pairs of data points, NLLSPP defines nonlocal or local similarity weight coefficients for pairwise data points on the basis of their distances. The nonlocal similarity weight coefficients force two nonadjacent data points to be projected far apart from each other. The local similarity weight coefficients force two neighboring data points to be projected near each other. In this way, nonlocal and local structures of the data set are naturally preserved and highlighted in a lower-dimensional space. Because of this advantage, NLLSPP is more powerful than principal component analysis (PCA) and locality preserving projections (LPP) in extracting important data characteristics. A process monitoring method is developed based on the NLLSPP algorithm. Its advantages are illustrated by a case study on the Tennessee Eastman process. The results indicate that the NLLSPP-based method has better monitoring performance that the PCA-based and LPP-based methods.

1. Introduction

Industrial processes are putting increasing demands on process safety, production efficiency and product quality. Process monitoring is able to guarantee the safe operation of the process and improve the product quality. Process monitoring aims to detect process faults that are caused by abnormal disturbances, improper operations, and so on. It provides the operators a chance to take measures to avoid casualty and equipment damage as well as reduce economic loss. The data-driven process monitoring, also termed as multivariate statistical process monitoring (MSPM), has been widely studied over the past few decades [1–5]. MSPM methods are easy to implement, because they mainly rely on process data and rarely need process knowledge. Because of this data-driven characteristic, MSPM methods are very suitable for complex industrial processes. To remove the redundant information and to extract important characteristics from process data, MSPM methods often reduce the dimension of process data using dimensionality reduction techniques. In this way, the important data information is preserved in a reduced subspace, while the redundant information is included in a residual subspace. Then, two monitoring statistics [6] are defined in the reduced subspace and the residual subspace respectively to monitor data variations caused by process faults.

Over the last few decades, various MSPM methods have been proposed on the basis of conventional dimensionality reduction techniques, such as principal component analysis (PCA) [1], partial least squares (PLS) [7], locality preserving projections (LPP) [8], Fisher discriminant analysis (FDA) [9], and so on. PCA and its variants, such as kernel PCA (KPCA) [10] and dynamic PCA (DPCA) [11], have been widely used in MSPM. PCA seeks a set of orthogonal projection axes to maximize the variance (i.e., the global scatter) of all projected data points [12]. It can preserve the global Euclidean structure of the data set. However, PCA may lead to the loss of local neighborhood information, because it ignores neighborhood relationships between data points [13]. This drawback also makes PCA easy to be affected by outliers and noise. Different from PCA, LPP aims to preserve the local neighborhood structure of the data set by minimizing the distances (i.e., the local scatter) between projections of neighboring data points [14]. However, LPP does not take into account the nonlocal (or global) data structure. Because of this drawback, LPP cannot guarantee that two nonadjacent data points are projected far apart from each other, and thus all data points may be projected into a very narrow region [13]. As a result, the global geometric structure of the data set may be destroyed, which leads to the loss of global data information. Therefore, neither PCA nor LPP can completely extract the important information from the data set. The PCA-based and LPP-based MSPM methods may have limited monitoring ability.

http://dx.doi.org/10.1016/j.chemolab.2016.07.014
0169-7439/

In recent years, to overcome shortcomings of PCA and LPP, some new linear dimensionality reduction algorithms have been proposed and applied for process monitoring [15–20]. These algorithms can simultaneously preserve global and local data structures. For example, Zhang et al. [15] proposed the global-local structure analysis (GLSA) by combining the objective functions of PCA and LPP with a weight coefficient. Yu [17] proposed the local and global principal component analysis (LGPCA) by maximizing the ratio between objective functions of PCA and LPP. Luo [18] developed a global-local preserving projections (GLPP) algorithm, which builds a unified framework for global structure preservation and local structure preservation. Although these methods have better performance than PCA and LPP, they inherit some drawbacks from LPP. Similar to LPP, they use "Heat kernel" or "binary" weight coefficients to weight the neighborhood relationships between data points [14]. An auxiliary parameter is required to define the "Heat kernel" weight coefficients [14], and its value has great effect on the dimensionality reduction performance. However, it is difficult to choose an optimal value for this parameter. The "binary" weight coefficients are very simple and contain no parameter [14], while they cannot reflect differences between different pairs of data points. These two drawbacks may degrade the dimensionality reduction performance, and thus decrease the process monitoring ability.

In this paper, a new linear dimensionality reduction algorithm, which is named as nonlocal and local structure preserving projection (NLLSPP), is proposed and applied for process monitoring. In order to simultaneously preserve nonlocal and local structures of the data set, NLLSPP uses nonlocal and local similarity weight coefficients to control the distances between projected data points. Unlike LPP, similarity weight coefficients in NLLSPP are defined based on the distances between data points in a parameterless form, and meanwhile they take into account the differences between different pairs of data points. The nonlocal similarity weight coefficients force the nonadjacent data points to be projected far apart from each other. On the contrary, the local similarity weight coefficients force the neighboring data points to be projected near each other. The tradeoff between nonlocal structure preservation and local structure preservation is automatically adjusted by the nonlocal and local similarity weight coefficients. As compared to GLSA, LGPCA and GLPP, NLLSPP is easier to implement and has more stable performance, because it only needs one parameter for defining neighborhood relationships. Then, a NLLSPP-based process monitoring method is proposed for fault detection. Its effectiveness and advantages are illustrated by a case study on the Tennessee Eastman process.

2. Nonlocal and local structure preserving projection (NLLSPP)

2.1. Basic idea

Important characteristics of a data set are normally reflected in two aspects: nonlocal (or global) geometrical structure and local geometrical structure. The nonlocal geometrical structure controls the exterior shape of the data set. The local geometrical structure represents the internal neighborhood relationships between data points. Both nonlocal and local geometrical structures are crucial for characterizing a data set. PCA ignores the local data structure, while LPP neglects the nonlocal data structure. As a result, neither PCA nor LPP can fully unfold important characteristics of the data set into a lower-dimensional space. To overcome this drawback, a new linear dimensionality reduction algorithm is proposed, which is named as nonlocal and local structure preserving projection (NLLSPP). The aim of NLLSPP is to find an optimal lower-dimensional mapping that can preserve nonlocal and local data structures simultaneously.

To illustrate differences between PCA, LPP and NLLSPP, Fig. 1 compares the one-dimensional projections of PCA, LPP and NLLSPP for a 2-dimensional (2D) data set. Each data point in the data set is denoted by a solid circle. Projection axes of PCA, LPP and NLLSPP are marked

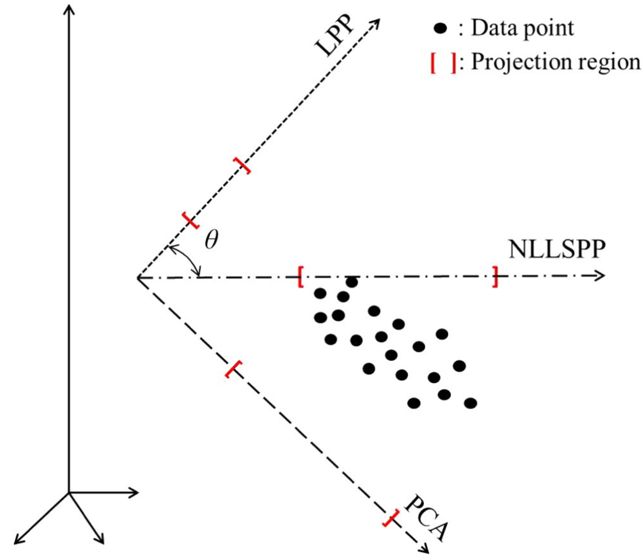

Fig. 1. Illustration of differences between PCA, LPP and NLLSPP.

by imaginary line, dotted line and dot dash line, respectively. Projection regions of PCA, LPP and NLLSPP are indicated by the red square bracket. In Fig. 1, the projection directions of PCA and LPP are almost mutually orthogonal. PCA projects data points along the direction that maximizes data variance. On the contrary, LPP projects data points along the direction that minimizes distances between neighboring data points. The projection axis of NLLSPP is between those of PCA and LPP, because it seeks to preserve both nonlocal and local data structures. The angle θ represents the tradeoff between nonlocal structure preservation and local structure preservation. In NLLSPP, the angle θ is determined by the nonlocal and local similarity weight coefficients that are defined in Section 2.2.

2.2. Objective function

Let $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\Re^{m\times n}$ be an m-dimensional training data set with n samples. It is projected into a lower-dimensional space as $\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n]\in\Re^{l\times n}$ $(l<m)$ via the linear mapping $\boldsymbol{y}_i=\boldsymbol{A}^T\boldsymbol{x}_i$ $(i=1,\ldots,n)$, where$\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_l]\in\Re^{m\times l}$ is a transformation matrix. NLLSPP aims to find an optimal $\boldsymbol{A}^*$ such that $\boldsymbol{Y}$ preserves both nonlocal and local structures of $\boldsymbol{X}$. Firstly, We consider the optimal mapping from the m-dimensional space to a line, i.e.,$\boldsymbol{y}=\boldsymbol{a}^T\boldsymbol{X}$, where $\boldsymbol{a}\in\Re^m$ denotes a transformation vector. To simultaneously preserve nonlocal and local structures of $\boldsymbol{X}$, the objective function of NLLSPP is constructed as

$$\max_{\boldsymbol{a}} J_{NLLSPP}(\boldsymbol{a}) = \max_{\boldsymbol{a}}[J_{NL}(\boldsymbol{a})-J_L(\boldsymbol{a})] \tag{1}$$

with

$$J_{NL}(\boldsymbol{a}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\notin N(\boldsymbol{x}_i)}\left(y_i-y_j\right)^2 W_{ij} \tag{2}$$

$$J_L(\boldsymbol{a}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\in N(\boldsymbol{x}_i)}\left(y_i-y_j\right)^2 W_{ij} \tag{3}$$

where $J_{NL}(\boldsymbol{a})$ and $J_L(\boldsymbol{a})$ are two sub-objective functions corresponding to the nonlocal structure preservation and the local structure preservation respectively, $J_{NL}(\boldsymbol{a})$ represents the nonlocal scatter of nonadjacent data points, $J_L(\boldsymbol{a})$ represents the local scatter of neighboring data points, $y_i=\boldsymbol{a}^T\boldsymbol{x}_i$ is the projection of $\boldsymbol{x}_i$, $N(\boldsymbol{x}_i)$ denotes the index set of neighbors of $\boldsymbol{x}_i$, and W_{ij} is the similarity weight coefficient. The $N(\boldsymbol{x}_i)$ can be defined by δ neighbors or k nearest neighbors. The δ neighbors is $N(\boldsymbol{x}_i)=$