Chemometrics and Intelligent Laboratory Systems 147 (2015) 58–65

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Data density-based fault detection and diagnosis with nonlinearities between variables and multimodal data distributions

Hiromasa Kaneko, Kimito Funatsu *

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:
Received 3 April 2015
Received in revised form 31 July 2015
Accepted 31 July 2015
Available online 7 August 2015

Keywords:
Process control
Fault detection
Fault diagnosis
Data density
One-class support vector machine

ABSTRACT

Multivariate statistical process control (MSPC) is an important means of monitoring multiple process variables and their interrelationships while controlling chemical and industrial plants efficiently and stably. To consider nonlinearities between process variables and multimodal data distributions, the data density can be used as an index for fault detection. Data domains with a low data density are considered abnormal states. However, after fault detection, faulty process variables cannot be diagnosed with an MSPC model based on the data density. Therefore, we have developed a new index to diagnose the process variables that contribute to process faults using a data density-based MSPC model. The proposed index uses the partial derivative of an MSPC model with respect to each process variable. We demonstrate the effectiveness of the proposed method using numerical simulation data, Tennessee Eastman process data, and real plant data analyses.

1. Introduction

For the safe and stable operation of industrial and chemical plants, it is necessary to monitor and control their operating conditions. Because of the huge amount of operating data in plants, data-based process control systems have received considerable attention in recent years. Controlling each process variable independently is inefficient, because there are many process variables that must be controlled. One practical solution is multivariate statistical process control (MSPC) [1,2], which monitors multiple process variables and their relationships simultaneously.

A major component of MSPC is principal component analysis (PCA) [3]-based process control [4]. After extracting the principal components (PCs) from multivariate normal data, the T^2 statistic is calculated as follows:

$$T^2 = \sum_{i=1}^{d} \frac{t_i^2}{s_i^2}, \tag{1}$$

where d is the number of PCs and s_i is the standard deviation of the ith PC. T^2 represents the magnitude of the principal variation in a training dataset. Errors in PCA models are monitored using the Q statistic, or squared prediction error, which is calculated as:

$$Q = \sum_{i=1}^{m} (x_i - \hat{x}_i)^2, \tag{2}$$

where m is the number of process variables and $\hat{x}_i$ is the process variable value estimated by the PCA model. A process fault is detected when one of the T^2 and Q statistics exceeds some threshold.

After fault detection with the T^2 and Q statistics, we must diagnose the cause of the fault. Therefore, the contribution of different process variables to the fault is calculated for each statistic [5,6]. We can use the indexes $ContT^2(\mathrm{x}_i)$ and $ContQ(\mathrm{x}_i)$ to represent the magnitude of the ith process variable x_i:

$$ContT^2(\mathrm{x}_i) = \mathbf{t}^{\mathrm{T}}\mathbf{S}^{-1}\left[x_i\mathbf{p}_i\left(\mathbf{P}^T\mathbf{P}\right)^{-1}\right]^{\mathrm{T}}, \tag{3}$$

$$ContQ(\mathrm{x}_i) = (x_i - \hat{x}_i)^2, \tag{4}$$

where $\mathbf{t}$ is a score vector for the training data, $\mathbf{S}$ is a variance–covariance matrix of a score matrix $\mathbf{T}$ for the training data, $\mathbf{p}_i$ is a loading vector corresponding to x_i, and $\mathbf{P}$ is a loading matrix. Process variables with large $ContT^2$ or $ContQ$ values are diagnosed to be the faulty variables. Zhao et al. combined PCA with rule-based reasoning approach to identify the root cause of faults in municipal solid waste incinerators [7].

* Corresponding author. Tel.: +81 3 5841 7751; fax: +81 3 5841 7771.
E-mail address: funatsu@chemsys.t.u-tokyo.ac.jp (K. Funatsu).

http://dx.doi.org/10.1016/j.chemolab.2015.07.016
0169-7439/

As well as PCA, partial least-squares (PLS) [8] and independent component analysis [9,10] have been applied to MSPC. When there is a nonlinear relationship between the process variables and the data distribution is multimodal, nonlinear MSPC methods [11] are more effective than linear ones. Jade et al. and Lee et al. applied kernel PCA (KPCA) [12], which is the extension of PCA to a nonlinear problem, to MSPC [13,14]. Cheng et al. added an adaptation mechanism to KPCA to give adaptive KPCA [15]. Godoy et al. applied kernel PLS to fault detection and developed a contribution plot for KPLS to identify faulty process variables [16]. A one-class support vector machine (OCSVM) can be applied to the domain description problem, enabling us to estimate domains in which the data density is high [17–19]. The k-nearest neighbors (k-NN) algorithm can also be applied to estimate the data density [20]. Data density estimation methods such as OCSVM and k-NN have been employed to distinguish process states and estimate prediction errors in soft sensors [21,22]. Such inferential models predict a difficult-to-measure process variable from easy-to-measure process variables [23].

Although MSPC models based on data density can handle nonlinearity between process variables and multimodal data distributions to detect fault situations, to the best of our knowledge, there is no way of diagnosing faulty process variables based on these nonlinear models after fault detection. Of course, Eqs. (3) and (4) cannot deal with nonlinearity between process variables and multimodal data distribution.

Therefore, we propose a method that diagnoses faulty variables using MSPC models and the data density. In this study, OCSVM and k-NN are used to estimate the data density. The absolute values of the partial derivative of a data density-based MSPC model with respect to each process variable provide an index of the contribution of each process variable to a fault. In the transition from a high-density domain to a low-density domain, process variables with higher index values will contribute more to the change from a normal process state to a process state in which the data density is low, i.e., an abnormal state. Once faulty process variables are diagnosed using our approach, rules or expert knowledge is required to diagnose the root cause of faults.

The effectiveness of the proposed method is demonstrated through two numerical simulation data analyses, the Tennessee Eastman process (TEP) [24], and an industrial membrane bioreactor (MBR) process.

2. Method

Because of its theoretical background and sparseness, OCSVM models can estimate the data density more accurately than k-NN models, although k-NN algorithms are generally simpler. Thus, we develop an expression for the partial derivative of an OCSVM model in this section, and give that for a k-NN model in Appendix A.

2.1. OCSVM

An OCSVM model f can be written as:

$$f\left(\mathbf{x}^{(i)}\right)=\phi\left(\mathbf{x}^{(i)}\right)\mathbf{w}-b \tag{5}$$

where $\mathbf{x}^{(i)}\in R^{1\times m}$ (m is the number of process variables or input variables; $\mathbf{X}$-variables) is an item of the (normally distributed) training data, ϕ is a nonlinear function, $\mathbf{w}$ is a weight vector, and b is a constant. If $f(\mathbf{x}^{(i)})>0$, $\mathbf{x}^{(i)}$ is within the high-density domain of the training data, and is diagnosed as a normal datum.

In OCSVM modeling, we aim to minimize

$$\frac{1}{2}\|\mathbf{w}\|^2+\frac{1}{\nu n}\sum_{i=1}^{n}\xi_i-b \tag{6}$$

subject to

$$\begin{aligned}&\phi\left(\mathbf{x}^{(i)}\right)\mathbf{w}\geq b-\xi_i,\\&\xi_i\geq 0\end{aligned} \tag{7}$$

where n is the number of training data. $\nu\in(0,1)$ is interpreted as the fraction of outliers in the training data, i.e., data for which $f(\mathbf{x}^{(i)})\leqq 0$. ν controls the size of the normal data domains. In this study, ν is set to 0.003 in reference to the three-sigma limit. ξ_i is a slack variable. Using the method of Lagrange multipliers, the output of a new datum $\mathbf{x}\in R^{1\times m}$ is given as follows:

$$f(\mathbf{x})=\sum_{i=1}^{n}\alpha_i K\left(\mathbf{x}^{(i)},\mathbf{x}\right)-b \tag{8}$$

where K is the kernel function given by:

$$K\left(\mathbf{x}^{(i)},\mathbf{x}\right)=\phi\left(\mathbf{x}^{(i)}\right)\phi(\mathbf{x})^{\mathrm{T}}. \tag{9}$$

Those $\mathbf{x}^{(i)}$ given by Eq. (6) for which $\alpha_i\neq 0$ are called support vectors. Only support vectors contribute to the OCSVM model f. In Eq. (8), b has the form:

$$b=\sum_{i=1}^{n}\alpha_i K\left(\mathbf{x}^{(i)},\mathbf{x}^{(\mathrm{S})}\right), \tag{10}$$

where $\mathbf{x}^{(\mathrm{S})}$ is a set of support vectors. The nonlinear function ϕ does not have to be explicitly defined, and the kernel function of Eq. (9) is sufficient to derive f. In this study, we use the Gaussian kernel:

$$K\left(\mathbf{x}^{(i)},\mathbf{x}^{(j)}\right)=\exp\left(-\gamma\left\|\mathbf{x}^{(i)}-\mathbf{x}^{(j)}\right\|^2\right), \tag{11}$$

where γ is a hyperparameter that can be optimized to maximize the variance in K using the training data [25,26]. LIVSVM [27] has been used as an optimization program to construct OCSVM models.

2.2. Partial derivative of an OCSVM model

Using Eq. (11), Eq. (8) can be transformed to:

$$\begin{aligned}f(\mathbf{x})&=\sum_{i=1}^{n}\alpha_i K\left(\mathbf{x}^{(i)},\mathbf{x}\right)-b.\\&=\sum_{i=1}^{n}\alpha_i\exp\left(-\gamma\left\|\mathbf{x}^{(i)}-\mathbf{x}\right\|^2\right)-b\\&=\sum_{i=1}^{n}\alpha_i\exp\left(-\gamma\sum_{k=1}^{m}\left(x_{i,k}-x_k\right)^2\right)-b\end{aligned} \tag{12}$$

The partial derivative of Eq. (12) with respect to the jth process variable x_j is then:

$$\frac{\partial f(\mathbf{x})}{\partial x_j}=2\gamma\sum_{i=1}^{n}\alpha_i\exp\left(-\gamma\sum_{k=1}^{m}\left(x_{i,k}-x_k\right)^2\right)\left(x_{i,j}-x_j\right). \tag{13}$$

Although we use a Gaussian kernel function in this paper, Eq. (8) can be partially differentiated for other differentiable kernel functions.

2.3. Diagnosis of faulty variables using a nonlinear MSPC model based on data density

When a datum is transiting from the domain of a normal state to that of an abnormal state, the change in data density for x_j should be large, because if the absolute value of Eq. (13) is large then x_j makes a large