Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 15740–15747

Quality-related Fault Detection Approaches Based on Data Preprocessing ⋆

Guang Wang * **Jianfang Jiao** ** **Shen Yin** ***

* *Bohai University, Jinzhou, 121013, China (e-mail: wangguang06041105@126.com).*
** *Bohai University, Jinzhou, 121013, China (e-mail: jiaojianfangheu@163.com).*
*** *Harbin Institute of Technology, Harbin, 150001, China (e-mail: shen.yin2011@googlemail.com)*

Abstract: This work focus on the issue of quality-related fault detection. A new idea of data-preprocessing is proposed with two specific instances are designed based on orthogonal signal correction (OSC) and orthogonal projections to latent structures (OPLS). Different from existing results, the new idea allows directly designing test statistics in the data subspaces obtained from data preprocessing without building any linear regression model like partial least squares (PLS). Benefit from such a direct feature, the designed new methods are more simple in engineering implementation and their performances are also more stable than conventional approaches. Simulation results on a widely used literature example and an industrial example demonstrate the effectiveness of the proposed new methods.

Keywords: Data-driven, fault detection, quality-related, data preprocessing.

1. INTRODUCTION

Quality-related fault detection (Zhou et al. (2010); Qin (2012); Zhang et al. (2015a)) is a new research subject of multivariate statistical process monitoring (MSPM) (Ding (2014); Chen et al. (2016, 2014); Ge et al. (2013); Yin et al. (2016a,b)) recently extracted from manufacturing industries. Traditionally, the main task of fault detection is to detect abnormal situations and alarm, so that supervising staff can take necessary remedial maintenance or even an emergency shutdown in a timely manner. However, long term industrial practices indicate that not all of the process faults will inevitably lead to the fluctuation of product quality. On the contrary, if ignore the alarms of the faults that have no effects on product quality, the unnecessary downtime and maintenance of factory can be significantly reduced, which finally brings considerable economic benefits (Peng et al. (2013b,a, 2015); Wang and Yin (2015); Wang et al. (2016); Jiao et al. (2016)). Therefore, quality-related fault detection approaches aim at a more targeted classification of the process faults, i.e. classifying the faults into the category of affecting product quality and the category of not affecting.

In actual industrial applications, abnormalities that seriously affect product quality should be paid more attention. The most direct way is to monitor quality variables space. However, quality variables are usually hard to be measured online or sampled with a significant time delay (Hao et al. (2014); Hu et al. (2014)). Since obtaining an accurate measurement result is expensive or even impossible, it is reasonable to model quality variables and process variables, then use the model to guide the implementation of fault detection scheme. For such a purpose, the MSPM methods are typically limited to PLS based methodologies (MacGregor et al. (1994)). Based on PLS model, many successful monitoring approaches have been developed (Qin (2012); Yin et al. (2012)). Recently, Li et al. (2010) revealed the geometric nature of PLS for process monitoring. Based on Li's result, Zhou et al. (2010) first analyzed the inherent flaw of PLS for quality-related fault detection and proposed a total PLS (T-PLS) model with a more detailed decomposition for process variables matrix. Zhou's work opened up the studies of quality-related fault detection. Subsequently, similar linear results emerged in the recent literatures (Qin and Zheng (2013); Yin et al. (2015); Peng et al. (2015); Wang et al. (2016); Shardt et al. (2015); Zhang et al. (2015a)). Relevant nonlinear research also attracts much attention and existing achievements can be found in recent literatures, for example, based on kernel partial least squares (KPLS) model Peng et al. (2013a) proposed the first nonlinear quality-related fault detection method by extending T-PLS to a nonlinear case, called total KPLS (T-KPLS). Then, Mori and Yu (2014) proposed a multi-way non-Gaussian latent subspace projection approach for nonlinear batch process. Soon later, Zhang et al. (2015b) extended the linear method of (Qin and Zheng (2013)) to a nonlinear version also based on KPLS model. Very recently, Jia and Zhang (2016) used KPLS model and SVD to realize another nonlinear method.

In fact, all the above methods can be attributed to the class of regression-model-based approaches. The composition of such methods usually involves three procedures, i.e. (i) construct a regression model like PLS or KPLS which builds the relationship between input and output

⋆ This work was supported by National Natural Science Foundations of China (No. 61503039, No. 61503040)

10.1016/j.ifacol.2017.08.2305

matrices; (ii) decompose the input matrix into several parts according to the correlation between the input and output matrices; and (iii) design appropriate test statistics in the decomposed parts to realize the aim of quality-related fault detection. In a nutshell, the core of quality-related method is dividing process variables space according to the correlation between process and quality variables spaces. Such being the case, data preprocessing technologies can be also a potentially viable way to achieve the aim of quality-related fault detection since their basic function is to remove the part uncorrelated with output from input. For such a purpose, this paper will propose a data-preprocessing-based idea for the same purpose as the regression-model-based ones. The major advancement of the new idea is that it directly filters the process variables space into two subspaces by data-preprocessing without a regression model. In the framework of the new idea, two specific instances are designed based on OSC (Wold et al. (1998)) and OPLS (Trygg and Wold (2002)), respectively. Thanks to their direct designing steps, the new methods are more simple in engineering implementation, and what's more, they are more stable than the conventional approaches.

Sec. 2 first gives the mathematical description of quality-related fault detection. Sec. 3 proposes two specific fault detection methods based on OSC and OPLS with theoretical proofs. In Sec. 4, simulations on a widely used literature example and an industrial example will be carried out to verify the effectiveness of the proposed two approaches. Finally, conclusions are drawn in Sec. 5.

2. MATHEMATICAL DESCRIPTION OF QUALITY-RELATED FAULT DETECTION

It is necessary to make some instructions for the considered industrial process and the processed data before discussion. Assume that, the given industrial process totally contains m process variables and l quality variables, the off-line measurements of which are recorded into observation samples $x_{obs}(k)$ and $y_{obs}(k)$, respectively, i.e.,

$$x_{obs}(k) = [x_{obs,1}, x_{obs,2}, \cdots, x_{obs,m}]^T \in \mathbb{R}^m \tag{1}$$

$$y_{obs}(k) = [y_{obs,1}, y_{obs,2}, \cdots, y_{obs,l}]^T \in \mathbb{R}^l \tag{2}$$

where $k = 1, 2, \cdots, N$. All the observation samples satisfy normal distributions and $N \gg m > l \geq 1$.

As a necessary step, all the observation samples should be normalized as follows $(i = 1, 2, \cdots m; j = 1, 2, \cdots l)$:

$$\bar{x}_{obs,i} = \frac{1}{N}\sum_{k=1}^{N} x_{obs,i}(k) \tag{3}$$

$$\sigma^2_{x,obs,i} = \frac{1}{N}\sum_{k=1}^{N} (x_{obs,i}(k) - \bar{x}_{obs,i})^2 \tag{4}$$

$$\bar{y}_{obs,j} = \frac{1}{N}\sum_{k=1}^{N} y_{obs,j}(k) \tag{5}$$

$$\sigma^2_{y,obs,j} = \frac{1}{N}\sum_{k=1}^{N} (y_{obs,j}(k) - \bar{y}_{obs,j})^2 \tag{6}$$

then, all the normalized values are recorded into the new samples $x(k)$ and $y(k)$,

$$x(k) = \left[\frac{x_{obs,1}(k) - \bar{x}_{obs,1}}{\sigma_{x,obs,1}}, \cdots, \frac{x_{obs,m}(k) - \bar{x}_{obs,m}}{\sigma_{x,obs,m}}\right]^T \tag{7}$$

$$y(k) = \left[\frac{y_{obs,1}(k) - \bar{y}_{obs,1}}{\sigma_{y,obs,1}}, \cdots, \frac{y_{obs,l}(k) - \bar{y}_{obs,l}}{\sigma_{y,obs,l}}\right]^T \tag{8}$$

next, the new samples $x(k)$ and $y(k)$ are composed into the sample matrices X and Y,

$$X = [x(1), x(2), \cdots, x(N)]^T \in \mathbb{R}^{N\times m} \tag{9}$$

$$Y = [y(1), y(2), \cdots, y(N)]^T \in \mathbb{R}^{N\times l} \tag{10}$$

For each online sample $x_{on} = [x_{on,1}, x_{on,2}, \cdots, x_{on,m}]^T \in \mathbb{R}^m$, it also needs to be normalized

$$x_{new} = \left[\frac{x_{on,1} - \bar{x}_{obs,1}}{\sigma_{x,obs,1}}, \cdots, \frac{x_{on,m}(k) - \bar{x}_{obs,m}}{\sigma_{x,obs,m}}\right]^T \tag{11}$$

Due to the fact that quality variables cannot be measured online, the process sample x_{new} is the only way to monitor process faults. However, the monitoring result of x_{new} cannot reflect whether current fault affects quality variables or not, because it has not builded the correlation between process variables space and quality variables space. In other words, if the process variables space is decomposed according to such correlation, the decomposed subspaces will have different correlations with the quality variables space. Accordingly, a quality-related fault detection method can be realized by designing appropriate statistics in these subspaces. That is, quality-related fault detection methods decompose the process variables space into the following forms:

$$x_{new} = C_1^{qr}x_{new} + \cdots + C_\nu^{qr}x_{new} + C_1^{qur}x_{new} + \cdots + C_\vartheta^{qur}x_{new} \tag{12}$$

where $C_i^{qr}, i = 1, 2, \cdots, \nu$ are ν projection matrices which project x_{new} onto the subspaces that are highly correlated with quality variables, while $C_i^{qur}, j = 1, 2, \cdots, \vartheta$ are ϑ projection matrices which project x_{new} onto the subspaces that are uncorrelated with quality variables, and

$$\sum_{i=1}^{\nu} C_i^{qr} + \sum_{j=1}^{\vartheta} C_j^{qur} = I_m \tag{13}$$

where I_m is an identity matrix with m-dimensional. Let,

$$x_{new} = \sum_{i=1}^{\nu} \hat{x}_{i,new} + \sum_{j=1}^{\vartheta} \tilde{x}_{j,new} \tag{14}$$

$$\hat{x}_{i,new} = C_i^{qr}x_{new} \in S_i^{qr}, i = 1, 2, \cdots, \nu \tag{15}$$

$$\tilde{x}_{j,new} = C_i^{qur}x_{new} \in S_j^{qur}, j = 1, 2, \cdots, \vartheta \tag{16}$$

Obviously, $S_i^{qr}, i = 1, 2, \cdots, \nu$ and $S_j^{qur}, j = 1, 2, \cdots, \vartheta$ constitute the entire process variables space, and faults happened in the former subspaces have effect on quality variables while faults happened in the latter subspaces have no effect on quality variables.

3. DATA PREPROCESSING BASED FAULT DETECTION APPROACHES

3.1 Orthogonal signal correction based fault detection approach

Usually, the prediction matrix X contains a large number of components that irrelevant with the response matrix Y.