Control Engineering Practice 67 (2017) 43–51

Contents lists available at ScienceDirect

Control Engineering Practice

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

A novel data-based quality-related fault diagnosis scheme for fault detection and root cause diagnosis with application to hot strip mill process

Liang Ma, Jie Dong, Kaixiang Peng *, Kai Zhang

Key Laboratory for Advanced Control of Iron and Steel Process of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, PR China

ARTICLE INFO

Keywords:
Quality-related
Root cause diagnosis
Hot strip mill process
Modified canonical variable analysis
Generalized reconstruction based contribution

ABSTRACT

In this paper, a new technology or solution of quality-related fault diagnosis is provided for hot strip mill process (HSMP). Different from traditional data-based fault diagnosis methods, the alternative approach is focused more on root cause diagnosis. The new scheme addresses the quality-related fault detection with the developed modified canonical variable analysis (MCVA) model, then the advantage of original generalized reconstruction based contribution (GRBC) is followed to identify the faulty variables. Meanwhile, a new transfer entropy (TE)-based causality analysis method is proposed for root cause diagnosis of quality-related faults. Finally, the whole proposed framework is practiced with real HSMP data, and the results demonstrate the usage and effectiveness of these approaches.

1. Introduction

To meet increasing market demands for multi-species, multi-specifications as well as high-quality products, modern industrial processes, such as HSMP, have become more and more complex and integrated. Associated with this trend, once an abnormality occurs somewhere in a plant, it may propagate to the whole HSMP or specific control loops by means of information and/or material flow pathways (Chioua et al., 2016; Choudhury, 2011; Yin, Liu, & Hou, 2016). The presence of abnormality may impact the overall process performance and the final products' quality, such as the thickness, flatness, width and so on. Thus, in order to maintain high-efficiency of the operation and ensure stability of the product quality, real-time fault detection, identification and accurate fault location are quite desired.

Traditionally, for most engineers, model-based methods act as basic tools to design and carry out some monitoring activities (Aouaouda, Chadli, Shi, & Karimi, 2015; Chibani, Chadli, Shi, & Braiek, 2016; Ding, 2013; Dong, Wang, & Gao, 2013; Gertler, 1998; He, Wang, Ji, & Zhou, 2011; Isermann, 2006; Li, Chadli, Ding, Qiu, & Yang, 2017; Yin, Zhu, Qiu, & Gao, 2016; Youssef, Chadli, Karimi, & Wang, 2017). Whereas, like HSMP, from a physical standpoint, the deformation of the thickness can be affected by the rolling force, the temperature, bending force and other physical properties that depend on the specific steel type. Therefore, to some extent, it is difficult to construct a precise model for HSMP, which results in the model-based methods cannot perform well and become invalid. In comparison, data-based methods, thanks to their simple forms and fewer requirements on the design and engineering efforts, have become more and more popular both in industry and academia domains nowadays (Ge, Song, & Gao, 2013; Kano & Nakagawa, 2008; Qin, 2003, 2012; Tidriri, Chatti, Verron, & Tiplica, 2016; Yin, Ding, Xie, & Luo, 2014; Yin, Li, Gao, & Kaynak, 2015; Yin, Wang, & Hao, 2016).

In actual HSMP, collected variables can be divided into process variables and quality variables in general. For process variables, such as gap, rolling force or bending force, can be measured online. However, measurements of quality variables like thickness or flatness should be realized after the production process is over. In practice, not all of the faults in process variables will influence the products' quality, and some may change the surrounding environment. Hence, identifying the covariances or correlations model between process variables and quality variables so as to monitor quality-related faults is very important in the industrial processes. For such a purpose, projection to latent structures (PLS)-based modeling methods have been significantly well-known in the quality-related fault detection for HSMP (Ding, Yin, Peng, Hao, & Shen, 2013; Peng, Zhang, Li, & Zhou, 2013; Peng, Zhang, You, & Dong, 2015a, b; Zhang, Hao, Chen, Ding, & Peng, 2015). However, several singular value decompositions (SVDs) are involved behind those algorithms while processing very large-scale industrial data. Canonical variable analysis (CVA), in contrast, is more efficient than PLS-based

* Corresponding author.
E-mail address: kaixiang@ustb.edu.cn (K. Peng).

http://dx.doi.org/10.1016/j.conengprac.2017.07.005
Received 7 December 2016; Received in revised form 14 July 2017; Accepted 14 July 2017
0967-0661/

methods, because it only needs one-step SVD (Hardoon, Szedmak, & Shawe-Taylor, 2004; Peng, Zhang, Dong, & Yang, 2014). Therefore, many pioneers have used CVA-based methods for fault detection in the past few years (Chen, Ding, Zhang, Li, & Hu, 2016; Chiang, Russell, & Braatz, 2000; Ding, 2014; Juricek, Seborg, & Larimore, 2004; Odiowei & Cao, 2009), which have largely improved monitoring performance for dynamic processes characterized by strong auto-correlated and cross-correlated variables.

Once a fault is alerted, all the above methods are capable of generating contribution plots to identify the major fault effect variables efficiently without prior process knowledge (MacGregor, Jaeckle, Kiparissides, & Koutoudi, 1994). Nevertheless, the root cause of faults cannot be located exactly, because these approaches are based on correlation other than causality among faulty variables (Li, Qin, & Yuan, 2016). As a result, some reasonable root cause diagnosis methods by capturing causality among different process variables have been developed in recent years.

Roughly speaking, the causality capture methods can be divided into two general groups: knowledge-based and data-based. Knowledge-based approaches are based on connectivity or causality that use operation mechanism, piping & instrument diagram (P&ID), process flow diagram (PFD) or expert knowledge (Duan, Chen, Shah, & Yang, 2014). A drawback is that these methods cannot provide any information on the level of interactions among variables (Yang, Duan, Shah, & Chen, 2014). By contrast, the data-based approaches are able to produce a quantitative model due to their abilities to measure to what extent the time series corresponding to specific variables influence each other, which have been widely used for investigating the causal interactions among process variables in the form of time series (Landman, Kortela, Sun, & Jömsö-Jounela, 2014). Among these data-based methods, cross-correlation function (CCF) (Bauer & Thornhill, 2008) and Granger causality (GC) (Landman et al., 2014; Yuan & Qin, 2014) are classical means. CCF is practical to be estimated from two time series and its results are arranged in a causality matrix, the GC method can identify the cause–effect relationship among process variables and capture the root cause of the plant oscillations that degrades the prediction performance, while it cannot explain whether the calculated GC comes from abnormal operations or not (Mori, Mahalec, & Yu, 2014). Alternatively, TE is utilized to quantify the size and direction of information flow for both linear and nonlinear relationships without any process knowledge, which has been successfully applied to various applications for estimating causal dependencies between time series (Bauer, Cox, Caveness, & Downs, 2007; Duan et al., 2014; Landman & Jömsö-Jounela, 2016; Lee et al., 2012; Schreiber, 2000; Yu & Yang, 2015). Nevertheless, the calculation of TE index may be a heavy burden if the number of sample is very large or the network topology is very complex. So adopting a suitable fault identification tool in advance to reduce the number of faulty variables as much as possible is a good choice.

Motivated by those observations, considering the logic and integrity of quality-related fault diagnosis, we will address three topics in this work: (1) developing a new MCVA-based quality-related fault detection method, and (2) presenting a GRBC-based quality-related fault identification method, and (3) proposing a TE-based method for root cause diagnosis of quality-related faults. Our major goal is to provide a high-efficiency tool to root cause diagnosis of quality-related faults for HSMP by combining process monitoring and causality analysis methods.

The rest of this paper is organized as follows. In Section 2, a novel MCVA-based quality-related fault detection approach is presented. After that, Section 3 is dedicated to the root cause diagnosis of quality-related faults. Then, the proposed scheme is implemented on the real HSMP in Section 4. In the end, the concluding remarks are presented in Section 5.

2. MCVA-based quality-related fault detection for dynamic processes

2.1. Conventional CVA statistical method

CVA is a classical dimensionality reduction tool that is optimal in terms of maximizing a correlation statistic between two or even more data sets. Suppose that the dynamic relationships between process variables and quality variables under consideration are modeled as a linear time-invariant (LTI) and with white process as well as measurement noise, it can be described by

$$\mathbf{x}(k+1)=\mathbf{A}\mathbf{x}(k)+\mathbf{B}\mathbf{u}(k)+\mathbf{w}(k) \tag{1}$$

$$\mathbf{y}(k)=\mathbf{C}\mathbf{x}(k)+\mathbf{D}\mathbf{u}(k)+\mathbf{v}(k) \tag{2}$$

where $\mathbf{x}\in\mathcal{R}^n$ represents a state vector, $\mathbf{u}\in\mathcal{R}^l$ and $\mathbf{y}\in\mathcal{R}^m$ are process variables and quality variables vectors, and $\mathbf{w}\in\mathcal{R}^n$ and $\mathbf{v}\in\mathcal{R}^m$ are zero-mean, Gaussian distributed white noises which are statistically independent of process variables. The system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are unknown constant matrices with appropriate dimensions.

Based on the above model (1) and (2), the dependence of the past vector $\mathbf{p}$ and the future vector $\mathbf{f}$ is investigated, which are traditionally defined as follows:

$$\mathbf{p}_k=[\mathbf{y}_{k-1}^{\mathrm{T}},\ldots,\mathbf{y}_{k-l'}^{\mathrm{T}},\mathbf{u}_{k-1}^{\mathrm{T}},\ldots,\mathbf{u}_{k-l'}^{\mathrm{T}}]^{\mathrm{T}} \tag{3}$$

$$\mathbf{f}_k=[\mathbf{y}_k^{\mathrm{T}},\mathbf{y}_{k+1}^{\mathrm{T}},\ldots,\mathbf{y}_{k+f'}^{\mathrm{T}}]^{\mathrm{T}} \tag{4}$$

where l' and f' are the numbers of lags. In general, $l'=f'$ is predefined, and the process order n is determined by Akaike information criterion (AIC) or cross-validation procedures, and selection based on the eigenvalues of the Hankel matrix (Ding, 2014; Larimore, 1996; Odiowei & Cao, 2009).

CVA seeks to find linear combinations of the future observations that correlate the most with the past observations, this correlation can be expressed as:

$$\rho_{J,L}=\frac{\mathbf{J}^{\mathrm{T}}\boldsymbol{\Sigma}_{pf}\mathbf{L}}{\sqrt{\mathbf{J}^{\mathrm{T}}\boldsymbol{\Sigma}_{pp}\mathbf{J}}\sqrt{\mathbf{L}^{\mathrm{T}}\boldsymbol{\Sigma}_{ff}\mathbf{L}}} \tag{5}$$

where $\mathbf{J}$ and $\mathbf{L}$ are transformation matrices, $\boldsymbol{\Sigma}_{pp}=\mathrm{E}(\mathbf{p}^{\mathrm{T}}\mathbf{p})$, $\boldsymbol{\Sigma}_{ff}=\mathrm{E}(\mathbf{f}^{\mathrm{T}}\mathbf{f})$ and $\boldsymbol{\Sigma}_{pf}=\mathrm{E}(\mathbf{p}^{\mathrm{T}}\mathbf{f})$.

This is equivalent to the following optimization problem:

$$\max_{J,L}=\mathbf{J}^{\mathrm{T}}\boldsymbol{\Sigma}_{pf}\mathbf{L}+\lambda_p(\mathbf{I}_p-\mathbf{J}^{\mathrm{T}}\boldsymbol{\Sigma}_{pp}\mathbf{J})+\lambda_f(\mathbf{I}_f-\mathbf{L}^{\mathrm{T}}\boldsymbol{\Sigma}_{ff}\mathbf{L}) \tag{6}$$

where $\mathbf{I}_p$ and $\mathbf{I}_f$ are identity matrices of appropriate dimensions, λ_p and λ_f are Lagrangian multipliers. The solution is given by SVD:

$$\boldsymbol{\Sigma}_{pp}^{-1/2}\boldsymbol{\Sigma}_{pf}\boldsymbol{\Sigma}_{ff}^{-1/2}=\hat{\mathbf{J}}\boldsymbol{\Sigma}\hat{\mathbf{L}}^{\mathrm{T}} \tag{7}$$

where $\mathbf{J}=\boldsymbol{\Sigma}_{pp}^{-1/2}\hat{\mathbf{J}}$, $\mathbf{L}=\boldsymbol{\Sigma}_{ff}^{-1/2}\hat{\mathbf{L}}$, and the main diagonal elements of $\boldsymbol{\Sigma}$ contains the correlation coefficients.

2.2. MCVA-based quality-related fault detection method

In this paper, for our application, inspired by DPCA- and DPLS- based methods (Ku, Storer, & Georgakis, 1995; Zhang, Shardt, Chen, Ding, & Peng, 2015), we extend the conventional CVA-based multivariate statistical method to address quality-related fault detection issues in stable and dynamic processes.

In order to maximize a more general measure of dependency and describe the correlation structure between variables completely, the Pearson's correlation coefficient in Eq. (5) will be replaced by mutual information (Yin, 2004). Then, the mutual information between $\mathbf{J}^{\mathrm{T}}\mathbf{p}$ and $\mathbf{L}^{\mathrm{T}}\mathbf{f}$ holds:

$$\mathbf{M}(\mathbf{J},\mathbf{L})=\mathrm{E}\left[\log\frac{p(\mathbf{J}^{\mathrm{T}}\mathbf{p},\mathbf{L}^{\mathrm{T}}\mathbf{f})}{p(\mathbf{J}^{\mathrm{T}}\mathbf{p})p(\mathbf{L}^{\mathrm{T}}\mathbf{f})}\right] \tag{8}$$

where $p(\cdot,\cdot)$ and $p(\cdot)$ are the joint probability density function and the marginal density function, respectively.

We find $\mathbf{J}_i$ and $\mathbf{L}_i$, where $i\le\min(l,m)$ such that

$$\mathbf{M}_i=\mathbf{M}(\mathbf{J}_i,\mathbf{L}_i)=\max\mathbf{M}(\mathbf{J},\mathbf{L}) \tag{9}$$