Control Engineering Practice 74 (2018) 1–12

Contents lists available at ScienceDirect

Control Engineering Practice

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

An incipient fault detection approach via detrending and denoising

Zhangming He [a,b], Yuri A.W. Shardt [c], Dayi Wang [b], Bowen Hou [a], Haiyin Zhou [a], Jiongqi Wang [a,*]

[a] College of Liberal Arts and Sciences, National University of Defense Technology, Fuyuan 1, 410073 Changsha, PR China
[b] Beijing Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing 100094, PR China
[c] Department of Chemical Engineering, University of Waterloo, 200 University Ave West, Waterloo, ON N2L 3G1, Canada

A R T I C L E I N F O

Keywords:
Incipient fault
Fault detection
Detrending
Denoising
Data-driven

A B S T R A C T

An incipient fault tends to be buried by either the process trend or the measurement noise. Fault–trend ratio (FTR) and fault–noise ratio (FNR) are two main factors that impact the detection performance. An incipient fault detection approach is proposed in this paper based on the detrending and denoising techniques. There are three main phases in this approach. First, to increase FTR, a detrending algorithm is implemented. The fault detection rate can be significantly enhanced, when the normal trend is eliminated from the testing residual. Second, to increase FNR, a denoising algorithm is realized. The residual obtained from this algorithm can avoid the incipient fault being buried by the widely oscillating noise. Therefore the fault detection performance can be further improved. Third, the new detection statistic is composed based on the two algorithms. The approach is applied to a simulated process, a satellite attitude control system process, and the Tennessee Eastman process. The comparison results show that the proposed method outperforms the traditional Hotelling method in detecting incipient faults.

1. Introduction

In modern industry, fault detection and isolation (FDI) is very important to enhance the system reliability, to prevent serious system performance deterioration, and to ensure optimal process operation (Shardt, Hao, & Ding, 2015; Shardt et al., 2012). Current FDI methods can be conventionally divided into two categories, qualitative methods and quantitative ones (Chen, Ding, Zhang, Li, & Hu, 2016; Chen, Zhang, Ding, Shardt, & Hu, 2016). The latter can be further divided into model-based methods and data-driven ones. With the development of sensors and databases, the amount of data available has grown quickly. Big data are characterized by 5 Vs: volume, variety, velocity, variability, and veracity. Data-driven methods become more and more important in the tasks of FDI. Artificial intelligent (AI) methods, *e.g.*, artificial neural networks (ANN) (Dong, Xiao, Liang, & Liu, 2008; Wang, 2003), support vector machines (SVMs) (Zhang, Zhou, Guo, Zou, & Huang, 2012), and fuzzy rough sets (Dong et al., 2008; Németh, Laboncz, Kiss, & Csépes, 2010), can also be regarded as data-driven methods, because the models in the AI methods are often trained by the monitored training data.

Residual generation is the key step of FDI. Residuals used for constructing the fault detection statistics are generated from either the model parameters or the monitored normal data. On the one hand, if the parameters of a input/output (I/O) model are known, then model-based residual generators are the preferred ones (Ding, 2013; Yin & Zhu, 2015), *e.g.*, fault detection filter (FDF), diagnostic observer (DO) (O'Reilly, 1983) and parity space (PS) (He et al., 2018; Patton & Chen, 1991). On the other hand, if the parameters of the I/O model are unknown, then data-driven residual generation techniques can be used instead, *e.g.*, principal component analysis (PCA) (Abdi & Williams, 2010; Hou et al., 2017; Wang, He, Zhou, Li, & Zhou, 2017), partial least square (PLS) (He, Zhou, Wang, Chen et al., 2015; Tobias et al., 1995), canonical correlation analysis (CCA) (Chen, Ding, Peng, Yang, & Gui, 2018; He, Zhou, Wang, & Zhai, 2015; Thompson, 2005), and identification of autoregressive models with external inputs (ARX) (Ding, 2014; Dong, Verhaegen, & Gustafsson, 2012; Ljung, 1998).

Incipient fault diagnosis has received considerable attention recently (Youssef, Delpha, & Diallo, 2016). If an incipient fault is neglected in its early stage, it may cause a disaster, *e.g.*, a small leak of nitrocellulose led to the Tianjin port explosion in 2015 in China (Sun, 2015), a tiny aging problem caused the explosion of the Fukushima nuclear power plant after an earthquake in 2011 in Japan (Tsubokura, Gilmour, Takahashi, Oikawa, & Kanazawa, 2012), a small fault in the wheel rims led to a German high speed train to go out of control in 1998 in Eschede (Oestern, Huels, Quirini, & Pohlemann, 2000), and a small deformation

* Corresponding author.
E-mail address: hezhangming@nudt.edu.cn (J. Wang).

https://doi.org/10.1016/j.conengprac.2018.02.005
Received 14 September 2017; Received in revised form 5 February 2018; Accepted 8 February 2018

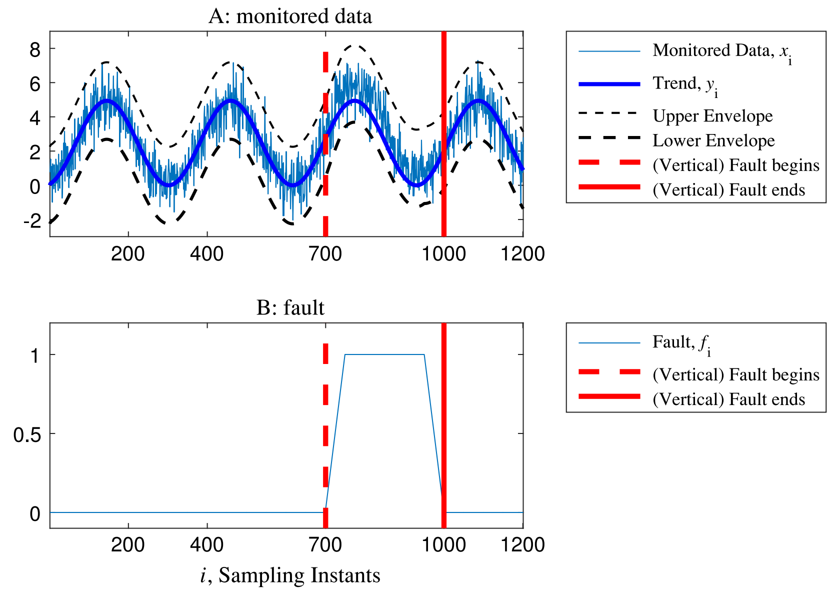

Fig. 1. The relationship between the trend, the noise, and the monitored data.

caused the gas leak that crashed the space shuttle *Challenger* in 1986 (Kramer, 1992). Incipient fault diagnosis in the modern nonlinear, non-stationary industrial processes is an important problem. The traditional fault diagnosis methods are mainly used to diagnose the faults with significant symptoms. An incipient fault usually cause a small shift in the observation variables, thus the diagnosis performance of the traditional methods are not feasible if they are used for diagnosing incipient faults without any preprocessing. Most current incipient fault diagnosis methods are based on signal processing techniques, *e.g.*, local approach based on the likelihood ratio (LR) (Cheng, 2000; Zhang, Basseville, & Benveniste, 1994), multivariate statistical or probability methods (Harmouche, Delpha, & Diallo, 2014, 2015; Youssef et al., 2016), wavelet transform methods (Carneiro, Da Silva, & Upadhyaya, 2008; Huo, Zhang, Francq, Shu, & Huang, 2017), Hilbert–Huang transform (HHT), empirical mode decomposition (EMD) (Yan & Lu, 2014), SVMs (Kang, Kim, & Kim, 2015; Liu, Yang, Zhang, Wang, & Chen, 2016; Long, Xian, Li, & Wang, 2014; Namdari & Jazayeri-Rad, 2014).

There are two common features of the existing methods:

(1) There is a lack of a strict definition of an incipient fault. Usually an incipient fault is considered to be a fault with small magnitude. In fact, an incipient fault should be defined based on the relative magnitude of the background information, *i.e.*, the trend and the noise. Generally speaking, a fault is incipient if its magnitude, compared with the range of the trend and the variance of the noise, is small.

(2) Most incipient fault diagnosis methods are applied for processes where the monitored data are periodical and oscillate around a constant mean value. Usually stationary trends are shown in the data of the monitored processes, *e.g.*, bearings (Huo et al., 2017; Kang et al., 2015; Li, Xu, Liang, & Huang, 2017; Liu et al., 2016), motors (Menacer, d Nai t Said, Benakcha, & Drid, 2014) and wind turbines (Sun, Zi, & He, 2014).

There are two natural questions:

(1) How to define and model an incipient fault? What are the main factors impacting the detection performance of incipient faults?

(2) How to improve the detection performance of incipient faults if the monitored data have a nonstationary trend and have widely oscillating noise?

Questions above will be answered in Sections 2 and 3, respectively, and then an incipient fault detection approach is developed to enhance the fault detection rate (FDR).

Notations. The transpose, the inverse, the Moore–Penrose pseudo-inverse, and the Frobenius-norm of matrix $\boldsymbol{A}$ are respectively denoted as $\boldsymbol{A}^{\mathrm{T}}$, $\boldsymbol{A}^{-1}$, $\boldsymbol{A}^{+}$, and $\|\boldsymbol{A}\|$. $\boldsymbol{A}_i$ denotes the ith column of $\boldsymbol{A}$. $\boldsymbol{0}$ denotes a matrix with proper rows and columns where all entries are zeros. $\mathrm{cond}(\boldsymbol{A})$ is the condition number of matrix $\boldsymbol{A}$, *i.e.*, the ratio of the maximal singular value to the minimal one. $P\{A\}$ is the probability of the event A. $\boldsymbol{r} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ denotes that $\boldsymbol{r}$ is a normal random vector with zero mean and covariance $\boldsymbol{\Sigma}$. $\chi^2(m)$ is a chi-square distribution with m degree of freedom and $\mathrm{F}(m,n)$ is a F-distribution with m and n degrees of freedom. $\chi^2_\alpha(m)$ and $\mathrm{F}_\alpha(m,n)$ are the upper critical points for $\chi^2(m)$ and $\mathrm{F}(m,n)$, respectively, corresponding to the significance level, α.

2. Factors impacting incipient fault detection performance

2.1. Trend, noise and incipient fault

As shown in Subfigure A of Fig. 1, the monitored data between the 700th and the 1000th samples are combined with a sinusoid trend, widely oscillating noise, and an incipient fault. It is hard to even visually see the fault in Subfigure A. The thick blue curve in the middle represents the normal trend. The measured data oscillates between the upper and lower envelopes caused by the noise. The fault is incipient and buried by the trend and the noise, because its magnitude is relatively small compared with the range of the trend or the confidence interval width of the noise.

At the i^{th} sampling instant, let $\boldsymbol{x}_i \in \mathbb{R}^{n_x}$ be the monitored data, $\boldsymbol{y}_i \in \mathbb{R}^{n_x}$ be the trend of the working system and $\boldsymbol{e}_i \in \mathbb{R}^{n_x}$ be the environment noise. Ideally, $\boldsymbol{y}_i$ should be equal to $\boldsymbol{x}_i$. However, the monitored process will inevitably affected by some kinds of noise, *e.g.*, electronic noise, wind, vibrations, gravitational attraction, variations in temperature, and variations in humidity, depending on what is measured and on the sensitivity of the device. Assume that

$$\boldsymbol{x}_i = \boldsymbol{y}_i + \boldsymbol{e}_i, \boldsymbol{e}_i \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{1}$$

where $\boldsymbol{e}_i \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ means that the noise vectors at different instants are independent and identically distributed, with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

Aging, environment change, and abnormal operation may cause a fault to occur in the monitored system, *e.g.*, constant fault, stuck fault, or drift fault either in the sensors or in the actuators, then the monitored datum may also involve a faulty signal, $\boldsymbol{f}_i \in \mathbb{R}^{n_x}$. Under the faulty condition, Eq. (1) becomes

$$\boldsymbol{x}_i = \boldsymbol{y}_i + \boldsymbol{e}_i + \boldsymbol{f}_i. \tag{2}$$