International Journal of Information Technology and Business Management
29th December 2016. Vol.55 No.1

ISSN 2304-0777 www.jitbm.com

A DATA-DRIVEN APPROACH FOR FAULT DETECTION WITH UNCERTAINTY IN HISTORICAL MODES

Gong Chen[1], Yuan Wei[2], Sun Zhou[1,*]

[1]Department of Automation, Xiamen University, China

[2]Department of Business Administration, Huaqiao University, China

Email: schow@126.com

Abstract

Bayesian methods are a kind of data-driven methods developed in recent years and have played an important role in fault detection and diagnosis. Nevertheless, traditional Bayesian fault detection methods cannot deal with the case where some underlying modes are ambiguous in the historical data. To cope with this problem, a new method is presented in this paper. The modes with uncertainty in historical data are classified under a Bayesian framework by combing historical data, current evidence and prior knowledge. In order to improve the detection performance, weighted kernel density estimation is employed in likelihood estimation. The proposed method is tested with Tennessee Eastman (TE) process benchmark data and shows better performance compared to previous approaches.

Keywords: fault detection, data-driven, likelihood estimation

1 INTRODCUTION

Bayesian methods are a kind of data-driven methods developed in recent years and have played an important role in fault detection. Bayesian inference provides a fit way to solve the industrial process monitoring and fault detection in the presence of disturbances. It is one of the most widely used methods in fault detection. Bayesian detection method is capable of incorporating historical data and prior knowledge for fault detection and is proved to perform well in industrial applications.

Fault detection and diagnosis has been an increasingly significant issue to improve the efficiency, stability and security for diverse processes. Owing to the complexity of industry processes, it is complicated to model a real industrial system. An alternative is to use data-driven methods such as Bayesian methods. It provides a reliable way for dealing with uncertainty conditions and is prove to be effective in engineering detection.

An important branch of fault detection and diagnosis method is control-loop performance detection and diagnosis that was first proposed by Harris (1989), and was further developed by Huang, Shah and Kwok (1995). Huang and Shah (1999) put forward a linear quadratic regulator (LQG) benchmark. Then further development was reported on evaluating performance of other components such as instruments (Gonzalez, Huang, Xu, Espejo, 2012; Qin & Li, 2001) and valves (Choudhury, Jain, Shah, 2008; He, Wang, Pottmann, Qin, 2007), etc. In these

International Journal of Information Technology and Business Management
29th December 2016. Vol.55 No.1

ISSN 2304-0777 www.jitbm.com

works, different monitoring algorithms are used to detect corresponding single problem and. Nevertheless, a change in operating mode may influence more than one instrument, resulting in the possibility of wrong detection. While Bayesian methods have no such limitation and have the superiority to detect unspecific faults. Nevertheless, one difficulty with Bayesian methods is that there must be exact knowledge of the underlying mode in the historical data.

To deal with such difficulty where there is uncertainty about the underlying mode, some approaches (Gonzalez & Huang, 2013a, 2013b) were proposed including incomplete data method, direct probability approach, second-order approximation method, etc. The incomplete data method just simply ignores the ambiguous data and the remaining data is utilized to apply Bayesian fault detection method. Thus, the detection performance is affected because of too few data is used for training. When the proportion of ambiguous data is bigger, the detection result will become more unreliable. Direct probability approach probability boundaries are too large to make a detection. Regarding the second-order approximation method, it approximates the likelihood through Taylor series expansion to yield better estimation, but there is still useful information that is not taken into consideration by this method. In this paper, a new approach is proposed, through which the ambiguous modes are softly classified by combing historical data, current evidence and the prior knowledge under a Bayesian framework.

2 DATA-DRIVEN DETECTION METHOD WITH UNCERTAINTY IN HISTORICAL MODES

2.1 Problem formulation

Let us take control loop detection as example. A typical control loop consists of the following components: controller, actuator, process, and sensor. Each component may subject to some abnormalities which may influence control loop performance. We assume that all or some monitors are designed for the components in the control loop. Nevertheless, each of monitors is possible to be affected by turbulences. In addition, any component may influence the monitors designed for other components. An allocation of operating state of all the components in the control loop is called a mode denoted as m , m can take different values and its specific value is denoted by m_k . The allocation of monitor readings is called evidence, and is denoted as $e=\{\pi_1,\pi_2,\ldots\pi_S\}$, where π_i is the i th monitor's reading, and S is the sum of monitor. Historical data are obtained from the past data record where the mode of control loop and the monitor readings are also recorded. Each sample d^t at time t in the historical data set D consists of the evidence e^t and the underlying mode m^t , which can be denoted as $d^t=\{e^t,m^t\}$.

Given evidence e , the historical data set D , posterior probability can be expressed as

$$p(m\mid e,D)=\frac{p(e\mid m,D)\,p(m\mid D)}{p(e\mid D)} \tag{1}$$

Where $p(m\mid e,D)$ is posterior probability, which is what we want to obtain under the condition of knowing current evidence e and historical data D . $p(m\mid D)$ is prior probability of mode m . $p(e\mid D)$ is a constant, and can be calculated as $p(e\mid D)=\sum_m p(e\mid m,D)\,p(m\mid D)$.

2.2 Data-driven approach

Here we present a Bayesian approach so that the ambiguous modes can be softly classified. By marginalization over all possible likelihood parameters, the likelihood

$$p(e\mid m=m_k,D)=\int_\Omega p(e\mid \Theta_{m_k},m=m_k,D)\,f(\Theta_{m_k}\mid m=m_k,D)\,d\Theta_{m_k} \tag{2}$$

The set $\Theta_{m_k}=(\theta_{e_1,m_k},\theta_{e_2,m_k},\ldots,\theta_{e_N,m_k})$ and $\theta_{e_i,m_k}=p(e=e_i\mid m=m_k)$ are parameters for all available evidence e . N is the total number of the evidence. Ω contains all parameters under all possible modes. $f(\Theta_{m_k}\mid m=m_k,D)$ can be calculated through the method of Bayesian rules:

ISSN 2304-0777 www.jitbm.com

$$f(\Theta_{m_k} \mid m = m_k, D) = \frac{p(D \mid \Theta_{m_k}, m = m_k) f(\Theta_{m_k} \mid m = m_k)}{p(D \mid m = m_k)} \qquad (3)$$

Then we assume that $f(\Theta_{m_k} \mid m = m_k)$ subjects to Dirichlet distribution, we can obtain the following:

$$f(\Theta_{m_k} \mid m = m_k) = \frac{\Gamma(\sum_{q=1}^{N} \alpha_q^{m_k})}{\prod_{i=1}^{N} \Gamma(\alpha_q^{m_k})} \prod_{q=1}^{N} \theta_{q,m_k}^{\alpha_q^{m_k}-1}, \alpha_q^{m_k} > 0$$

(4)

$\alpha_q^{m_k}$ is the parameter of Dirichlet distribution indicating prior sample number under mode m_k. $\Gamma(x)$ is the gamma function.

The probability $p(D \mid \Theta_{m_k}, m = m_k) = \prod_{t=1}^{N_{m_k}} p(d_{m_k}^t \mid \Theta_{m_k}, m = m_k)$, where $d_{m_k}^t$ is the data point at time t with mode m_k. When $e^t = e_i$

$$p(d_{m_k}^t \mid \Theta_{m_k}, m = m_k) = \theta_{i|m_k} \qquad (5)$$

The conclusion $p(d_{m_k}^t \mid \Theta_{m_k}, m = m_k) = \prod_{i=1}^{N} \theta_{i|m_k}^{n_{i|m_k}}$ is made. Then we can calculate the value of the likelihood.

$$p(e = e_i \mid m = m_k, D) = \frac{n_{i|m_k} + a_{i|m_k}}{N_{m_k} + A_{m_k}} \qquad (6)$$

$n_{i|m_k}$ is the number of historical data point when $e = e_i$ and $m = m_k$. $a_{e_i|m_j}$ is the amount of prior samples that is distributed to evidence e_i with mode m_j; $N_{m_j} = \sum_i n_{e_i|m_j}$ and $A_{m_j} = \sum_i a_{e_i|m_j}$.

Assume there is an ambiguous mode M_K. Before we calculate the proportion parameter, we should first figure out the following parameters through (6):

$$\hat{\theta}_{d_i}\{\frac{m_j}{M_K}\} = \frac{(n_{e_i|m_j} + a_{e_i|m_j}) / (N_{m_j} + A_{m_j})}{\sum_{m_k \in m_Q} (n_{e_i|m_k} + a_{e_i|m_k}) / (N_{m_k} + A_{m_k})}$$

(7)

where d_i denotes a data point in historical data set and $d_i = (e_i, M_K)$, $n_{e_i|m_j}$ is the amount of historical samples when the evidence $e = e_i$, and mode $m = m_j$; $a_{e_i|m_j}$ is the amount of prior samples that is distributed to evidence e_i with mode m_j; m_Q is a set of all available modes. In addition, where $N_{m_j} = \sum_i n_{e_i|m_j}$ and $A_{m_j} = \sum_i a_{e_i|m_j}$.

Then normalize $\hat{\theta}_{d_i}\{m_j / M_K\}$ to obtain proportion parameters:

$$\theta_{d_i}\{\frac{m_j}{M_K}\} = \hat{\theta}_{d_i}\{\frac{m_j}{M_K}\} \Big/ \sum_{m \in M_K} \hat{\theta}_{d_i}\{\frac{m_j}{M_K}\} \qquad (8)$$

Summary the formulas (6) and (8), likelihood can be estimated as

$$p(e_i \mid m) = \frac{n(e_i \mid m) + \sum_{m \in M_k} (\theta_{d_i}\{\frac{m}{M_k}\} n(e_i \mid M_k)) + a_{e_i|m}}{n(m) + \sum_{d_i \in D} \sum_{m \in M_k} (\theta_{d_i}\{\frac{m}{M_k}\} n(e_i \mid M_k)) + A_m} \qquad (9)$$

Then we can obtain the final posterior probability through Bayesian method.

ISSN 2304-0777 www.jitbm.com

3 IMPROVE PERFORMANCE WITH WEIGHTED KDE

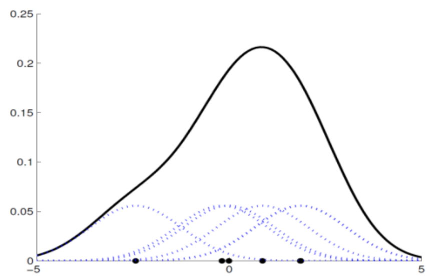

Figure1 Kernels summing to a kernel density estimate

In addition, to enhance detection we propose weighted kernel density estimation instead of classic histogram method for likelihood estimation.

The advantage of kernel density estimation is that it reflects the true shape of the data and can fully reflects the distributions regardless of the data as shown in Fig.1. In addition, a kernel density estimate can be obtained in a simple step. So, kernel density estimates can yield the same result even with a single data set. Nevertheless, it is in strong connection with the discrete method, the kernel density method subjects to the issue of dimensionality. When in high dimension, performance will depress.

Kernel density estimation (Gonzalez & Huang, 2014) performs precisely instead of histogram density estimation. Diverse modes correspond to diverse weights. Assume the weight of mode m is θ_m, the kernel density estimation function will be:

$$\hat{p}(x) = \frac{1}{Nh}\sum_{i=1}^{N}\theta_m K\left(\frac{x - x_i}{h}\right) \qquad (10)$$

where N is the total number of evidence, and h is a kernel density constant, $K(\cdot)$ is kernel function. x_i denotes the current evidence. $\hat{p}(x)$ is the estimation function.

4 SIMULATION WITH TE PROCESS

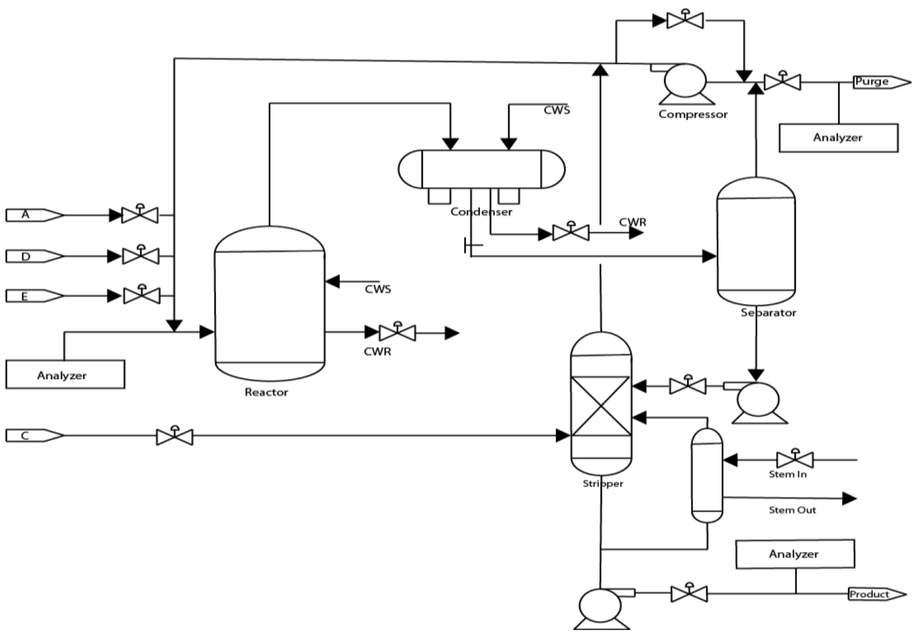

Figure 2 The Tennessee Eastman process (Ricker, 1996)

To verify the capability of the proposed method, it is applied to detection of Tennessee Eastman problem. The Tennessee Eastman Plant-wide Industrial Process Control Problem is called as TE problem (Downs & Vogel, 1993), which has been viewed as a criterion to assess the performances of various detection methods. Fig.2 shows a procedure of the TE process. The process comes into being two products from four reactants. Also, there is a byproduct and an inert. The reactions are:

International Journal of Information Technology and Business Management
29th December 2016. Vol.55 No.1

ISSN 2304-0777 www.jitbm.com

$$\begin{aligned} A(g) + C(g) + D(g) &\rightarrow G(liq), \text{Product1} \\ A(g) + C(g) + E(g) &\rightarrow H(liq), \text{Product2} \\ A(g) + E(g) &\rightarrow F(liq), Byproduct \\ 3D(g) &\rightarrow 2F(liq), Byproduct \end{aligned} \quad (11)$$

The gaseous reactants are sent to the reactor, in which the reactants reactor to come into being liquid products. The products run out of the reactor in the form of vapor as well as the unreacted reactants, and enter a cooler to be condensed. Then they go through a vapor/liquid separator. The components that are not condensed return to the reactor by a compressor. The condensed components are sent to the stripper to remove the rest reactants by reacting with reactant C. The products export the stripper at the bottom of the system. The inert and byproduct are purged from the system in the form of vapor.

In this paper, we implement our method with the data set used by Russell, Chiang, and Braatz (2000). (available at http://web.mit.edu/braatzgroup), where the TE process is controlled by the measure proposed by Lyman and Georgakis (1995). In the paper, three representative modes (IDV2: B composition, A/C ratio constant, IDV6:A feed loss,IDV8:A,B,C feed composition) were selected, and two measurement variables (XMEAS(4), XMEAS(8)) were used.

In this simulation, since both training data and test data have exact information about evidence and mode, we shall mask some of the data as ambiguous according to resemblance of data toward other modes. During the masking course, a parameter called likelihood ratio threshold is established, as a result, if another mode is similar enough, the data point will be classified as ambiguous. For example, assume the data comes from mode m_1.If there is a data point d^k, the likelihood ratio R between mode m_k and mode m_1 is large enough, then the mode m_k was added. As a result, the mode with d^k is assumed to be ambiguous. By adjusting the value of the threshold, certain amounts of data can be classified as ambiguous. After that, the aforementioned data sets are utilized in the proposed method.

$$R = \frac{p(d^k \mid m_k)}{p(d^k \mid m_1)} > Threshold \quad (12)$$

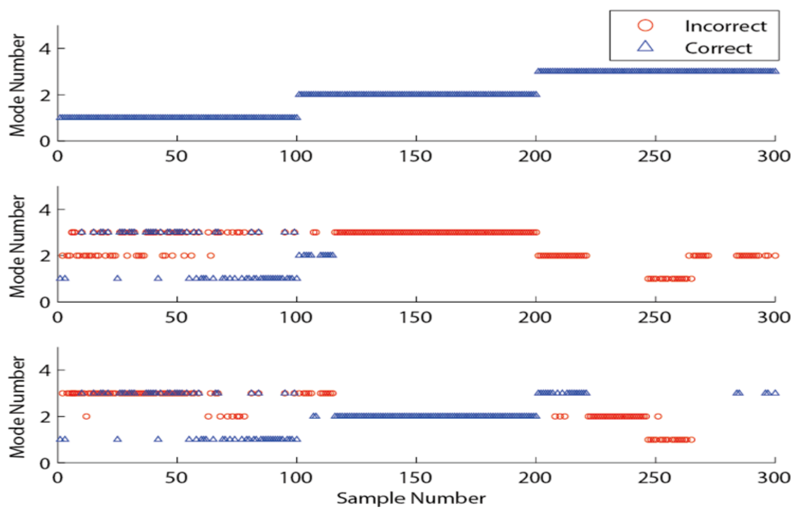

Figure 3 Detection result comparison

Here we apply the incomplete method and the proposed method to the fault detection problem of TE problem. The final posterior probability calculated by each of the methods is shown in Fig. 3. Horizontal coordinates of every graph in Fig.3 denotes all possible modes, while longitudinal coordinate denotes the possibility. The first three graphs are the detection result with incomplete data method, and the others are results by proposed method. One hundred data are taken from each mode to test the result.

ISSN 2304-0777 www.jitbm.com

When Mode=3, a misdetection result is obtained by incomplete data method. As a comparison, the proposed method performs better, with which we can obtain a precise result. Fig.3 also shows the detection result, from which we can briefly observe the determined modes for each data point. The red circle represents data points with wrong detected modes, while the blue triangles indicate correct detection results. Horizontal coordinates of each graph in Fig.4 denotes the number of the validating data points. Data points with numbers ranging from 1 to 100 are taken from mode=1; with numbers ranging from 101 to 200 are taken from mode=2; and the remaining 100 data points are under mode=3.The first graph of Fig.4 show the true mode of all the 300 data points. The second is the detection result using incomplete data method, and the third is obtained through proposed method. Obviously, the proposed method yields better detection result.

5 CONCLUSIONS

For fault detection problems where there is uncertainty about the underlying mode, a new approach is put forward under Bayesian framework to combine historical data, current evidence and prior knowledge together and to softly classify the data under ambiguous mode. Furthermore, weighted kernel density estimation instead of classic histogram method is proposed for likelihood estimation to enhance detection. The proposed Bayesian approach is tested on the fault detection problem of Tennessee Eastman (TE) process and shows better performance compared to typical previous methods.

References

[1] Choudhury, M., Jain, M., Shah, S. (2008). Stiction - definition, modeling, detection and quantification. Journal of Process Control, 18(3-4), 232-243.

[2] Downs, J.J., Vogel, E.F. (1993). A Plant-Wide Industrial Process Control Problem. Computers and Chemical Engineering, 17, 245-255.

[3] Gonzalez, R., Huang, B., Xu, F., Espejo, A. (2012). Dynamic Bayesian approach to gross error detection and compensation with application toward an oil sands process. Chemical Engineering Science, 67(1), 44-56.

[4] Gonzalez, R., Huang, B. (2013a). Control loop diagnosis with ambiguous historical operating modes: Part 1. A proportional parameterization approach. Journal of Process Control, 23, 585-597.

[5] Gonzalez, R., Huang, B. (2013b). Control loop diagnosis with ambiguous historical operating modes: Part 2, information synthesis based on proportional parameterization. Journal of Process Control, 23, 1441 -1454.

[6] Gonzalez, R., Huang, B. (2014). Control-loop diagnosis using continuous evidence through kernel density estimation. Journal of Process Control, 24, 640-651.

[7] Harris, T.J. (1989). Assessment of control loop performance. Canadian Journal of Chemical Engineering, 67(5), 856-861.

[8] He, Q., Wang, J., Pottmann, M., Qin, S. (2007). A curve fitting method for detecting valve stiction in oscillation control loops. Industrial and Engineering Chemistry Research, 46(13), 4549-4560.

[9] Huang, B., Shah, S., Kwok, K. (1995). On-line control performance monitoring of MIMO processes. American Control Conference, 2, 1250-1254.

[10] Huang, B., Shah, S. (1999). Performance Assessment of Control Loops, Springer.

[11] Lyman, P.R., Georgakis, C. (1995). Plant-wide control of the Tennessee Eastman problem. Computers and Chemical Engineering, 19, 321-331.

[12] Qi, F., Huang, B. (2010). A Bayesian Approach for Control Loop Diagnosis with Missing Data. AIChE Journal, 56(1), 179-195.

[13] Qin, S.J., Li, W. (2001). Detection and identification of faulty sensors in dynamic processes. AIChE Journal, 47(7), 1581-1593.

[14] Ricker, N. (1996). Decentralized control of the Tennessee Eastman challenge process. Journal of Process Control, 6(4), 205-221.

[15] Russell, E.L., Chiang, L.H., Braatz, R.D. (2000). Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes, Springer.