Journal of Process Control 54 (2017) 152–171

Contents lists available at ScienceDirect

Journal of Process Control

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

Data-driven sensor fault diagnosis systems for linear feedback control loops

Kai Wang [a], Junghui Chen [b,*], Zhihuan Song [a,*]

[a] State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, 310027, Zhejiang, China
[b] Department of Chemical Engineering, Chung Yuan Christian University, Chungli, Taoyuan, 32023, Taiwan, ROC

ARTICLE INFO

Article history:
Received 20 May 2016
Received in revised form 4 February 2017
Accepted 2 March 2017

Keywords:
Dynamic
Fault diagnosis
Feedback control
Residual
State-space model

ABSTRACT

This paper develops a sensor fault diagnosis (SFD) scheme for a multi-input and multi-output linear dynamic system under feedback control to identify different types of sensor faults (bias, drift and precision degradation), particularly for the incipient sensor faults. Feedback control, leading to fault propagation and disguised fault rectification, imposes the challenge on the data-driven SFD. With only available output data in closed loop, the proposed scheme comprises two stages of residual generation and residual evaluation. In the residual generation, a data-driven identification of the residual generator for the feedback control system is proposed. One class of parameters in the residual generator are estimated using process delays while another class of parameters describing the output dynamic are derived by the Bayes' formula. The means and variances control charts of online calculated residuals are made to judge the root cause. Two case studies are performed to illustrate the effectiveness of the proposed method.

1. Introduction

To ensure proper functionality of complex systems, advanced technologies for performance diagnosis and control have been widely incorporated into engineering designs to solve sophisticated, expensive and safe systems [1,2]. These performance diagnosis and control functionalities necessitate the use of an ever-increasing number of sophisticated sensors and measurement instruments to provide us with valuable information about operating conditions and critical quantities. Measurement instrument faults may result in improper action of the process control system and/or incorrect conclusion drawn by the process fault detection and diagnosis system, which can cause performance degradation or operation shutdown. This in turn necessitates the implementation of more reliable sensor fault detection and diagnosis methods [3–5].

The conventional engineering method for sensor validation checks and recalibrates the sensor periodically by following a set of predetermined procedures [5]. With the increasing number of interconnected subsystems and associated sensors, it has become less and less feasible and cost-effective to check all the sensors periodically. On the other hand, the hardware redundancy approach has been widely used in many complex and safe systems [6]. Although the method is relatively easy to implement and it can grant a high certainty in the detection and isolation of faulty sensors, the use of redundant sensors may not be always feasible due to the cost and space constraints. To avoid the use of redundant sensors, sensor fault detection and diagnosis in dynamic systems has received more and more attention over the last two decades, both in a research context and also in industries.

Most contributions in sensor fault diagnosis (SFD) rely on the analytical redundancy principle. This basic idea is to use accurate system models to capture the dynamics of the system as well as the sensors themselves. Based on a nominal model established in the fault-free conditions, residuals can then be generated as the difference between the actual sensor readings and the values estimated from the nominal model. If a fault occurs, the residual signal can be used to identify the malfunction. Several residual generation schemes have been well developed in literature [7–11]. The most frequently used analytical redundancy methods include diagnostic observer, Kalman filter and parity relations. These residual generators are designed based on a deterministic system model, such as the input-output model or the state space model with prior knowledge of model structures and parameters. Redundancy in the model-based design can help construct

* Corresponding authors.
E-mail addresses: jason@wavenet.cycu.edu.tw (J. Chen), songzhihuan@zju.edu.cn (Z. Song).

http://dx.doi.org/10.1016/j.jprocont.2017.03.001

the corresponding residual generator, making further fault isolation achievable. Nevertheless, as it is mentioned, the analytical redundancy principle needs an accurate analytical model of the system as well as the sensors, so prior deep understanding of their underlying physics is required. Inspired by the fact that the measurements of a sensor depict the dynamic characteristics of the system as well as those of the sensor itself, system identification techniques have been employed to construct an empirical model, such as state space models, auto-regressive-moving-average models and transfer functions, to identify the analytical relations among the measured variables in the dynamic system. Since the operating system is typically not permitted to operate in an open loop, the closed loop identification of the dynamics of the system as well as the sensors themselves is usually necessary, but compared with the open-loop identification, the closed loop identification is not easily successful because there are correlations between output noises and system inputs [12–17].

With the development of process instrumentation and data collection techniques, a large amount of data can be easily stored in the database. Multivariate variable analysis (MVA) approaches have been widely developed as alternatives in the last ten years such as principle component analysis (PCA) [3,4,18] or partial least squares (PLS) [19,20]. They are more flexible and rarely rely on the process knowledge. To cope with the process dynamics, several methods, including the dynamic recursive, the fast moving window and the multi-mode variants of MVA algorithms, have been developed [4,18,21–23]. Then based on the contribution of each measurement variable to the individual score of the multivariable latent models, the contribution plot technique [24] is used to identify the most likely root cause. However, there are several limitations when MVA approaches are applied. First, the collected data contain a large amount of data with the stationary behavior, but there is only a small number of data in the initial transient stage because most of the fluctuations would be eliminated by feedback control systems. Although the data in the initial transient stage often contain the critical information for diagnosis and control systems, it is still difficult to detect incipient sensor faults. This problem has seldom received attention in literature. Second, a sensor which is declared faulty may appear in four forms, namely: bias, drift, complete failure and precision degradation. Except for complete failure which means the sensor reading remains constant regardless of the changes in the actual value, the fault will be compensated by the controller actions. This will let the fault be masked because of the sensor located at the feedback loop. Moreover, in the multivariable control loop, there is often coupling. If there is coupling, the sensor fault effects can propagate to other feedback loops and they can significantly impact the overall process performance [25,26]. Therefore, large sensor fluctuations can respond to other faulty sensors in real measurements so that it is erroneously detected as a fault.

The process monitoring methods, including the analytical redundancy methods and data driven methods, consist of two main stages: off-line for setting control limits and on-line for testing. At the off-line stage, a residual generation based on system models is designed in analytical redundancy methods. Some statistics and confidence lines would be constructed using the normal data in data-driven methods. At the on-line stage, a new sample is applied to the model to calculate the residual or statistics for evaluating whether the sample exceeds its control limit. If the sample is out of the control limit, the sample is marked as a faulty event; sequentially the detection and further diagnosis of the faults are conducted.

An accurate model of the complex system that can predict the evolution of measured variables is difficult to obtain. From the viewpoint of the SFD performance, both the MVA approach and analytical principle models have advantages and drawbacks no matter whether they have statistical or mechanistic nature for closed loop sensor fault diagnosis. In process monitoring methods, the mechanistic methods and data driven methods, the precision of the models has great influence on the monitoring performance. In terms of mechanistic methods for process monitoring, there is no difference between the open loop and the closed loop because the accurate process model is known in advance. However, considering the data-driven methods, the collected data in the open loop and the closed loop may be differently informative. Because of the feedback control actions, the correlations occurring in the output variables come from the influence of the input variables, and their outputs may be transmitted to part of or all the input variables. Thus, there are correlations between the noises (process noises and measurements noises) and the process inputs. According to the system identification theories, the estimated process parameters would be biased if the input and the output data from the closed loop are directly applied. They will cause the significant degradation of the monitoring performance. However, most of the process monitoring problems in literature only considered the open-loop situation, but they did not mention the open-loop explicitly. Those methods have been systematically summarized in Ding's book [23].

Moreover, in the past, there was not much work on data driven based model approaches for closed loop sensor fault diagnosis; still, some of the approaches are promising. Gertler and Cao [27] proposed PCA-based fault diagnosis in the presence of control and dynamics to enhance analytical redundancy approaches. They concluded that PCA would be effective by changing the reference signals or controller parameters for fault isolation under feedback control. However, the changes are often not permitted in practice. McNabb and Qin [28] used one-order prediction errors of outputs to construct the feedback invariant subspace, which was said to preserve open loop sensor fault directions, but Wan and Ye [29] found that the fault direction of one-order predictive errors could not eliminate the closed loop's influence. They developed a residual covariance based method, but the sensor fault is limited only to the precision degradation. The faulty sensor reading is caused by different types of faults.

In order to address the above issues, a good method for the detection and diagnosis of a faulty sensor should have the following desirable characteristics. (1) It is easy to distinguish where a sensor fault occurs in the feedback control loop. (2) It is able to detect and diagnose an incipient faulty sensor promptly for the purpose of diagnosis and control in a real-time environment. (3) It is capable of detecting and diagnosing a faulty sensor even in the case when multiple faulty sensors occur at the same time and the faults come from bias, drift or precision degradation. In this paper, a residual based SFD scheme is proposed to solve the above problems. With the assumption that only closed loop data in the normal operation are available, the data driven approach constructs a residual generator based on the overall closed loop outputs. The vector of the residual generator has a special structure and each element of the residual vector is paired with exactly one of the faulty sensors. The remaining part of the paper is organized as follows. Section 2 gives the problem formulation of SFD in the MIMO feedback control system. The weak points will be explained when the conventional SFD methods for the open loop system is directly applied to the feedback control system. Then the detailed residual generator model is proposed in Section 3 Particularly, the estimations of the parameters of the residual generator for the overall closed loop system are clearly derived. Section 4 gives a detailed residual evaluation and diagnosis procedure. Illustrative examples are given in Section 5 to present the performance of the proposed method through two sets of benchmark data from a numerical example and a simulation column. The final section gives conclusions to the paper.