Annual Reviews in Control 36 (2012) 220–234

Contents lists available at SciVerse ScienceDirect

Annual Reviews in Control

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

Survey on data-driven industrial process monitoring and diagnosis

S. Joe Qin

Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089, USA
Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA

ARTICLE INFO

Article history:
Received 8 June 2012
Accepted 9 September 2012
Available online 26 October 2012

Keywords:
Statistical process monitoring
Quality monitoring
Fault detection
Fault diagnosis
Principal component analysis
Partial least squares
Machine learning

ABSTRACT

This paper provides a state-of-the-art review of the methods and applications of data-driven fault detection and diagnosis that have been developed over the last two decades. The scope of the problem is described with reference to the scale and complexity of industrial process operations, where multi-level hierarchical optimization and control are necessary for efficient operation, but are also prone to hard failure and soft operational faults that lead to economic losses. Commonly used multivariate statistical tools are introduced to characterize normal variations and detect abnormal changes. Further, diagnosis methods are surveyed and analyzed, with fault detectability and fault identifiability for rigorous analysis. Challenges, opportunities, and extensions are summarized with the intent to draw attention from the systems and control community and the process control community.

1. Introduction

Data-driven process monitoring or statistical process monitoring (SPM) applies multivariate statistics and machine learning methods to fault detection and diagnosis for industrial process operations and production results, which has become one of the most fruitful areas in research and practice over the last two decades. Based on methods from multivariate statistical analysis, SPM has found wide applications in various industrial processes, including chemicals, polymers, microelectronics manufacturing, iron and steel, pharmaceutical processes, and power distribution networks. The SPM tasks typically include: (i) fault detection; (ii) fault identification or diagnosis; (iii) fault reconstruction that estimates the fault magnitudes and fault-free values; and (iv) product quality monitoring and control. Due to the data-based nature of the SPM methods, it is relatively easy to apply to real processes of rather large scale comparing to other methods based on systems theory or rigorous process models.

The use of multivariate statistics for abnormal situation detection has been studied intensively in the area of multivariate quality control (MQC) (Jackson, 1991). Typically the Hotelling's T^2 statistic and the Q-statistic, which is also known as the squared prediction error (SPE), are used for the detection of an out-of-control situation. These two statistics, calculated based on a model from principal component analysis (PCA), give an overall account of abnormal situations in a complementary manner. The MQC literature, however, mainly focuses on the monitoring of quality variables and the detection of a quality problem. However, it is difficult for MQC to pinpoint to process variables that lead to the quality problems.

The work of Kresta, MacGregor, and Marlin (1989, 1991) Kresta, MacGregor, and Marlin (1991) and that of Wise, Veltkamp, Ricker, and Kowalski (1988, 1991) are among the first to apply multivariate methods to process variables in addition to quality variables. PCA and partial least squares (PLS) are the methods of choice in characterizing the normal situations from data. Although these papers use virtually the same two statistics for fault detection as those used in MQC, later process monitoring work extends the use of multivariate statistics for fault reconstruction and diagnosis, (Chiang, Russell, & Braatz, 2001; Dunia & Qin, 1998a, 1998b; Dunia, Qin, Edgar, & McAvoy, 1996; Gertler, Li, Huang, & McAvoy, 1999; MacGregor, Jaeckle, Kiparissides, & Koutoudi, 1994; Miller, Swanson, & Heckler, 1993; Qin et al., 1999; Raich & Cinar, 1996; Stork & Kowalski, 1999; Wise et al., 1988; Yue & Qin, 1998). Extensions to modeling dynamic process data are available in, for example, (Ku, Storer, & Georgakis, 1995; Qin & Li, 2001). Nonlinear extensions can be achieved by principal curve neural networks (Dong & McAvoy, 1996), auto-associative neural networks (Kramer, 1991), and kernel PCA (KPCA) (Cho, Lee, Choi, Lee, & Lee, 2005). Other variations of modeling methods, such as independent component analysis (ICA) (Kano, Tanaka, Hasebe, Hashimoto, & Ohno, 2003; Lee, Qin, & Lee, 2006), aim to extract non-Gaussian latent components. These methods can be categorized as *latent variables methods* (LVMs), which use latent variable models that differ from traditional input-output models or causal models (Yoon & MacGregor, 2000).

E-mail address: sqin@usc.edu

1367-5788/$ - see front matter
http://dx.doi.org/10.1016/j.arcontrol.2012.09.004

Data-driven fault detection and diagnosis are widely accepted in industrial practice due to the data-driven and multivariate characteristics. Early adoption is reported at Dupont (Piovoso, Kosanovich, & Yuk, 1992). Other successes are found in several industries including semiconductor manufacturing (Yue, Qin, Markle, Nauert, & Gatto, 2000; Cherry & Qin, 2006), the steel industry (Miletic, Quinn, Dudzic, Vaculik, & Champagne, 2004; Zhang & Dudzic, 2006; Kano et al., 2008), and the chemical industry (Kosanovich, Dahl, & Piovoso, 1996). The process industries are characterized by large-scale complex operations that are well instrumented with a multi-level control hierarchy, making it a suitable place to apply the data-driven methodologies.

The industrial process problem considered for the data-driven fault detection and diagnosis can be illustrated in Fig. 1. The process is the centerpiece of the focus whose efficient operation to make good quality products is desired. Disturbances and variabilities in raw material and operation conditions are the obstacles to overcome in order to make good quality products. Process instrumentation and control are in place to maintain normal and efficient operations. To assure quality, additional quality measurements are usually available, but they come with much less frequency (as indicated by the dotted arrows) and a long time delay since quality variables are difficult and expensive to measure. As with other engineering systems, the added instrumentation and control, along with the manufacturing processes, are all prone to abnormal conditions (or soft faults) and failures (or hard faults), leading to quality and efficiency losses. The task of data-driven process monitoring is to detect such an abnormal situation and diagnose the root-cause early. The monitoring and diagnosis tasks depicted in Fig. 1 include

- Faults in sensors and actuators are the main focus of the traditional fault detection and isolation (FDI) research as a field in control engineering. Robust methods are available to be insensitive to model errors and process disturbances to certain extent, see, for example, (Gertler, 1998; Isermann, 2006; Ding, 2008).
- Control performance monitoring and assessment are concerned with the degradation of control performance with the changing process or disturbance characteristics. See, for example, (Huang, Shah, & Kwok, 1997; Qin, 1998; Harris, Seppala, & Desborough, 1999).
- PCA type of process monitoring, including nonlinear PCA and ICA, extracts latent variables from process measurements under feedback control and monitors changes in the condition of the processes, disturbances, sensors, and actuators, which may or may not lead to poor product quality. This task overlaps with the sensor and actuator FDI methods develop in control engineering.
- PLS type of monitoring, including nonlinear PLS, uses quality data to guide the decomposition of the process data and extract latent variables that are most relevant to the product quality. Diagnosis capability in terms of quality relevance is enhanced and false alarm rates are reduced due to the use of quality data (Li, Qin, & Zhou, 2010; Li, Alcala, Qin, & Zhou, 2011). This task is usually not handled in the areas of traditional fault detection and isolation (Gertler, 1998; Isermann, 2006; Ding, 2008).
- Quality control is mainly concerned with the final control of the quality of the products, with no effort to correlate them to process conditions reflected in process measurements.

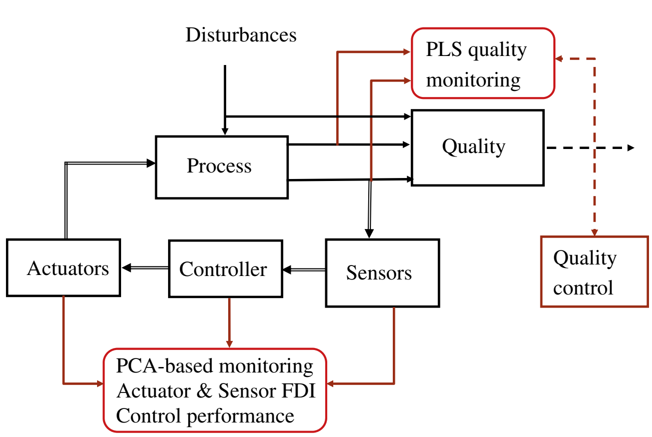

Fig. 1. Process and quality monitoring problem considered in data-driven fault detection and diagnosis.

The multivariate SPM methods based on PCA and PLS models offer a practical approach for fault detection and diagnosis. While fault detection is accomplished by directly applying statistics used in MQC, fault diagnosis or classification is made possible by the use of contribution plots (Miller et al., 1993; Wise & Ricker, 1991; Kourti & MacGregor, 1994; Tong & Crowe, 1995; Alcala & Qin, 2009), fault identification via reconstruction (Dunia et al., 1996; Dunia & Qin, 1998b), and reconstruction-based contributions (RBC) (Alcala & Qin, 2009). Work by Gertler et al. (1999) describes an isolation-enhanced PCA approach which uses a bank of PCA models for fault identification. A structured residuals approach with maximized sensitivity for fault diagnosis is proposed in Qin et al. (1999).

Although the statistical process monitoring methods have been under development for about two decades, interest in both academia and practice is going strong from the process industries (Qin, 2009) and the fault detection community (Ding, 2012). Early accounts of reviews and monographs in the areas have focused on the essential aspects of multivariate statistics for process monitoring that brought significant impact to the area in terms of methodology development and industrial practice (MacGregor & Koutoudi, 1995; Wise & Gallagher, 1996; Chiang et al., 2001; Qin, 2003; Cinar, Palazoglu, & Kayihan, 2007; Kano et al., 2008; Yao & Gao, 2009). The objective of this paper is to provide a survey of the recently developed process monitoring methods for fault detection, reconstruction, identification, and diagnosis. Fault diagnosability analysis is included for various diagnosis methods. The scope of the monitoring problem is described with reference to the scale and complexity of industrial process operations, where multi-level hierarchical optimization and control are necessary for efficient operation, but are also prone to hard failure and soft operational faults that lead to economic losses. Recent development and extensions such as nonlinear kernel methods and dynamic approaches are reviewed. The connection of the data-driven methods to the model based approaches developed in the control community is addressed. Conclusions and future opportunities are given at the end of the paper.

2. Process and fault modeling

Two different methodologies are available in process monitoring and fault diagnosis. One is to build models for all interested fault cases using first-principles-driven or data-driven models, the other is to build a model for the normal case only and use it to detect faults that deviate from the normal case. It is obvious that the former methodology requires much more modeling effort than the latter one. For this reason most statistical process monitoring methods rely on the use of normal process data to build normal-case process models. These models range from PCA, PLS, and other variants. PCA models are dominantly used to extract process