Computers and Chemical Engineering 35 (2011) 319–328

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Finding a trade-off between observability and economics in the fault detection of chemical processes

M. Bin Shams, H. Budman*, T. Duever

Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 5 November 2009
Received in revised form 9 March 2010
Accepted 2 April 2010
Available online 10 April 2010

Keywords:
Observability
Fault detection
Average run length
PCA
Economic significance

ABSTRACT

This paper presents a methodology to quantitatively gauge the potential economical loss due to unobserved faults when standard statistical monitoring charts are used. It is shown that in closed loop operation, a shorter time for detection may result from retuning the controller at the expense of higher product variability. Accordingly, an optimization approach is proposed for finding a trade-off between the economic losses resulting from lack of detection and losses resulting from higher product variability. In order to account for faults with different frequency contents, the method is applied in the frequency domain. The proposed optimization based methodology is later validated in the time domain.

1. Introduction

The need for efficient and profitable operation in chemical industries requires the use of efficient process monitoring strategies. Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003, emphasized that the petrochemical industry looses over $20 billion per year due to inappropriate reaction to abnormal process behavior. Thus, faults have a serious impact on process economy, product quality, safety, productivity and pollution level. A fault may be defined as a deviation of at least one variable from an acceptable level (Isermann, 2006). The survey papers (e.g. Gertler, 1988; Himmelblau, 1978; Isermann, 1984; Willsky, 1976) provide a summary of early work in this area and (Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003) provide a more recent account. Most of the available fault detection algorithms involve comparing the observed behavior of the process to the corresponding output of a reference model which may be mechanistic, empirical or semi empirical (Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003). If the fault is observable, the fault detection scheme will generate fault symptom patterns which in turn are fed to the fault diagnosis scheme to determine the root cause of the observed abnormal behavior.

A fault diagnostic system is composed of a detection algorithm followed by a diagnosis scheme. An observable fault is defined as one that can be detected or observed from the chosen set of measurement variables in spite of the background noise. Lack of observability will result in a suboptimal operation due to the presence of an undetected fault.

When data is collected from a process while a fault is occurring, the application of a given statistical model to these data, either univariate or multivariate, is supposed to indicate the presence of the fault. If the statistical model fails to provide indication of the fault this may signify that the specific fault cannot be observed with that particular model. The most common reasons for this lack of observability are as follows: (a) the measured process variables exhibit low signal to noise ratios and (b) the measured variables do not contain sufficient information regarding this fault and more representative variable(s) should be used for detection (Raghuraj, Bhushan, & Rengaswamy, 1999; Kourti, 2002). The latter reason is especially important when those variables used for detection are tightly controlled to satisfy quality requirements resulting in lack of information with respect to the fault detection scheme. Then, in order to detect a fault, it may be required to increase the variability, for example, by detuning the controller, so the fault can be observed.

On the other hand, detuning the controller causes deterioration of closed loop performance and possible loss of profit due to

* Corresponding author. Tel.: +1 519 888 4567x36980; fax: +1 519 746 4979.
E-mail address: hbudman@uwaterloo.ca (H. Budman).

0098-1354/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2010.04.006

higher product variability. Hence, there is a tradeoff between fast fault detection on the one hand and good control on the other. Most of the available fault detection systems, in particular data driven techniques, are implemented as a supplement to the available control system. Despite the significant amount of research in fault detection, the topic of the interaction between control and fault diagnosis has not been extensively studied in particular in the context of fault observability and fault distinguishability. Jacobson and Nett (1991) proposed a four parameter controller setup as a generalization of the two degrees of freedom controllers and Tyler and Morari (1994) reformulated the four degrees of freedom controller into a general framework for which tools from optimal and robust control were applied. The main conclusion of their studies was that when uncertain plants are used in synthesizing a model based controller, the control and diagnostic systems must be synthesized simultaneously. The main drawbacks of these approaches are: (a) they did not use standard fault diagnostic algorithms (e.g. exponential weighted moving average (EWMA), cumulative sum (CUSUM), principal component analysis (PCA), partial least square (PLS), etc.) and (b) they did not address the economic impact of unobservable faults. The focus of this work is to investigate the simultaneous design of controller and fault diagnosis scheme to enhance fault observability while mitigating through control the impact of unobserved faults.

This work addresses these topics as follows:

1. A tabular CUSUM and T^2-PCA based algorithms are used for detection for univariate and multivariate cases, respectively. Under low signal to noise ratio, it is shown that these algorithms require a certain period of time to detect certain classes of faults. Accordingly, the observability of the fault is related to its duration or alternatively to its frequency.
2. The tuning parameters of the closed loop controller are optimized to achieve an optimal tradeoff between economic losses that may result when high frequency faults (relative to a statistical monitoring chart) and closed-loop variability are experienced.

The paper is organized as follows. In Section 2, definitions and theoretical background are presented. The details of the algorithm and the models are given in Section 3. To illustrate the methodology, a simulation example based on an endothermic continuous stirred tank reactor is presented in Section 4. Analysis and discussion of the result are presented in Section 5 followed by conclusions.

2. Preliminaries and theoretical background

2.1. The tabular CUSUM

The cumulative sum chart is an efficient control chart that accumulates information collected over current and past samples. The tabular CUSUM is performed using the following two statistics

$$C_j^+ = \max(0, C_{j-1}^+ + x_i - (\mu_{in\,control} + K)) \quad (1)$$

$$C_j^- = \max(0, C_{j-1}^- + (\mu_{in\,control} + K) - x_i) \quad (2)$$

where C_0^- and C_0^+ are equal to zero. When either C_j^- or C_j^+ exceed a threshold H, the process is out of control. Based on an extensive study, Montgomery (1997) provides guidelines for the selection of K and H. In general, $K = k{\cdot}\sigma_{in\,control}$ and $H = h{\cdot}\sigma_{in\,control}$, where $\sigma_{in\,control}$ is the in control standard deviation. A common choice for k and h are $\delta/2$ and 5, respectively. The variable δ represents the expected shift in standard deviation units induced in the monitored variable.

2.2. Principal component analysis (PCA)

For a process with n measurement variables, one alternative is to use n univariate control charts to monitor the process. In order to simplify the presentation of information, a second alternative consists of using a principal component analysis (PCA) model to produce T^2 and Q charts for monitoring the n variables simultaneously. PCA involves the computation of loadings and scores using the covariance matrix of data $X \in R^{m\times n}$; where n is the number of variables and m is the total number of samples (MacGregor & Kourti, 1995). If the original variables are correlated, it is possible to summarize most of the variability present in the n variables space in terms of a lower p dimensional subspace ($p \ll n$). Here, p represents the number of the principal components. If only two principal components are found, two-dimensional score plots are used (i.e. T_1 versus T_2). For more than two principal components, Hotelling T^2 and Q statistics are usually used to monitor the process. The T^2 statistics based on the first p PCs is defined as

$$T^2 = \sum_{i=1}^{p} \frac{t_i^2}{\lambda_i} \quad (3)$$

where λ_i is the ith eigenvalue of the covariance matrix of the original data matrix. Confidence limits for T^2 at confidence level $(1-\alpha)$ are related to the F-distribution as follows:

$$T_{m,p}^2 = \frac{(m-1)p}{m-p} F_{p,m-p} \quad (4)$$

where $F_{m,m-p}$ is the upper $100\alpha\%$ critical point of the F-distribution with p and $(m-p)$ degrees of freedom. Monitoring the process variables by the T^2 values based on the first p principal components is not sufficient since this will only help detect whether or not the variation is within the plane defined by the first p principal components which generally captures steady state correlation. If a new event which was not present in the calibration data used to identify the reference model occurs, then additional principal components may become significant and the new observation vector x_i will move off the calibrated plane. Such new events can be detected by computing the squared prediction error or Q statistic. Let $x_i \in R^n$ denote the ith multivariate observation vector whose corresponding score is $t_i = x_i{\cdot}P$. The prediction from the PCA model for x_i is given by $\hat{x}_i = t_i \cdot P^T = x_i P \cdot P^T$. Then, the p dimensional error vector is given by $e_i = x_i - \hat{x}_i$ and the corresponding Q is defined as follows

$$Q = e_i \cdot e_i^T \quad (5)$$

Accordingly Q can be thought of as a measure of plant-model mismatch. The confidence limits for Q are given by Jackson (1991). This test suggests the existence of abnormal condition when $Q > Q_\alpha$, where Q_α is defined as follows

$$Q_\alpha = \Theta_1 \left(1 + \frac{c_\alpha \cdot h_0 \cdot \sqrt{2\cdot\Theta_2}}{\Theta_1} + \frac{\Theta_2 \cdot h_0(h_0-1)}{\Theta_1^2} \right)^{1/h_0} \quad (6)$$

$$\Theta_i = \sum_{j=p+1}^{n} \lambda_j^i; \quad \text{for } i = 1, 2, 3 \quad (7)$$

$$h_0 = 1 - \frac{2\cdot\Theta_1\cdot\Theta_3}{3\cdot\Theta_2^2} \quad (8)$$

c_α are the confidence limits for the $(1-\alpha)$ percentile in a standard normal distribution. These confidence limits are calculated based on the assumptions that that the measurements are time independent and multivariate normally distributed.