Journal of Loss Prevention in the Process Industries 40 (2016) 365–377

Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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

Amalgamation of anomaly-detection indices for enhanced process monitoring

Fouzi Harrou [a, *], Ying Sun [a], Sofiane Khadraoui [b]

[a] King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

[b] University of Sharjah, Department of Electrical and Computer Engineering, Sharjah, United Arab Emirates

ARTICLE INFO

Article history:
Received 14 August 2015
Received in revised form
4 December 2015
Accepted 26 January 2016
Available online 29 January 2016

Keywords:
Principal component analysis
Anomaly detection
Data-based approach
Control charts

ABSTRACT

Accurate and effective anomaly detection and diagnosis of modern industrial systems are crucial for ensuring reliability and safety and for maintaining desired product quality. Anomaly detection based on principal component analysis (PCA) has been studied intensively and largely applied to multivariate processes with highly cross-correlated process variables; however conventional PCA-based methods often fail to detect small or moderate anomalies. In this paper, the proposed approach integrates two popular process-monitoring detection tools, the conventional PCA-based monitoring indices Hotelling's T^2 and Q and the exponentially weighted moving average (EWMA). We develop two EWMA tools based on the Q and T^2 statistics, T^2-EWMA and Q-EWMA, to detect anomalies in the process mean. The performances of the proposed methods were compared with that of conventional PCA-based anomaly-detection methods by applying each method to two examples: a synthetic data set and experimental data collected from a flow heating system. The results clearly show the benefits and effectiveness of the proposed methods over conventional PCA-based methods.

1. Introduction

1.1. The state of the art

Identifying anomalies in modern automate process is central to process safety and maintaining product quality. Because anomalies in complex processes are common and some can be the source of serious degradations, their early detection is desirable. Accurate and effective anomaly detection and diagnosis of modern engineering systems are crucial for ensuring reliability and safety and for maintaining desired product quality; most importantly detection may prevent disasters in the cases of equipment failure. Early anomaly detection is favorable for reducing operational and maintenance costs and system down-time, while increasing levels of safety. The purpose of anomaly detection is to identify any event in the process behavior that is deviation from its nominal behavior (Isermann, 2006).

Early and accurate anomaly detection is important for increasing process safety and for limiting losses caused by abnormal and faulty states. The operation of a plant or a process has motivated the various anomaly detection and diagnosis methodologies proposed in the literature (Serpas et al., 2013; Venkatasubramanian et al., 2003a, c; Qin, 2012). These techniques can be broadly classified into three main categories: (i) data-based or model-free techniques, (ii) model-based techniques and (iii) knowledge-based techniques. Knowledge-based anomaly detection is usually a heuristic process (Venkatasubramanian et al., 2003a). Techniques in this category are mostly based on causal analysis, expert systems (Kim et al., 2005), possible cause and effect graphs (Wilcox and Himmelblau, 1994), failure modes and effects analysis (Wirth et al., 1996), hazard and operability analysis (Venkatasubramanian et al., 2003b) or Bayesian networks (Verron et al., 2008). These methods are best suited to small-scale systems with a small number of variables, and thus may not be appropriate for monitoring complex processes. Model-based monitoring methods compare process-measured variables with information obtained from a mathematical model, which is usually developed based on some fundamental understanding of the process under fault-free conditions (Gertler, 1998; Staroswiecki, 2001). This approach relies on the concept of analytical redundancy (Kinnaert, 2003). More specifically, after a model is developed, it is

* Corresponding author.
E-mail address: fouzi.harrou@kaust.edu.sa (F. Harrou).

http://dx.doi.org/10.1016/j.jlp.2016.01.024
0950-4230/

Nomenclature

T^2	Hotelling statistic
m	Number of variables
n	Number of observations
$\mathbf{X}\in\mathbb{R}^{n\times m}$	Data matrix
$\mathbf{X}_s\in\mathbb{R}^{n\times m}$	Autoscaled data matrix
l	Number of principal components
$\mathbf{T}\in\mathbb{R}^{n\times m}$	Score matrix
$\mathbf{P}\in\mathbb{R}^{m\times m}$	Eigenvector matrix
$\mathbf{E}$	Error matrix
$\widehat{\mathbf{X}}$	Approximated data matrix
$\Sigma\in\mathbb{R}^{m\times m}$	covariance matrix
Λ	Eigenvalue matrix of the PCA model
$\widehat{\mathbf{T}}\in\mathbb{R}^{n\times l}$	Matrix of l retained principal components
$\tilde{\mathbf{T}}\in\mathbb{R}^{n\times(m-l)}$	Matrix of $(m-l)$ ignored principal components
$\widehat{\mathbf{P}}\in\mathbb{R}^{m\times l}$	Matrix of l retained eigenvectors
$\tilde{\mathbf{P}}\in\mathbb{R}^{m\times(m-l)}$	Matrix of $(m-l)$ ignored eigenvectors
$p_i\in R^m$	ith eigenvector of $\mathbf{P}$
$t_i\in R^n$	column vector of $\mathbf{T}$
γ	EWMA smoothing constant
L	Width of the EWMA control limits
z_t	EWMA decision statistic

used to compute the process residuals, which are then evaluated to detect the presence of anomalies. The residuals, which are the difference between the measurements and the model prediction, can be used as an indicator of the presence or absence of anomalies in a monitored process (Kinnaert, 2003; Staroswiecki, 2001). When the monitored process is under normal operating conditions (no anomaly), the residual is zero or close to zero in cases of modeling uncertainties and measurement noise. When an anomaly occurs, the residuals deviate significantly from zero, indicating the presence of a new condition that is significantly distinguishable from the normal faultless working mode (Kinnaert, 2003; Harrou et al., 2014). The model-based anomaly-detection approaches include observer-based approaches (Kinnaert, 2003), parity-space approaches (Staroswiecki, 2001) and interval approaches (Benothman et al., 2007). Naturally, the effectiveness of these monitoring methods depends on the accuracy of the models used; however, deriving accurate models of monitored systems can be difficult, especially for complex processes. In addition, anomalies that have not been considered in the modeling stage may not be detected at all. Consequently, this problem, data-based anomaly-detection techniques are more commonly used.

Data-based monitoring methods, also known as process-history-based methods or black-box methods (Venkatasubramanian et al., 2003c), use the process data collected during normal operating conditions to build an empirical model that describes the nominal behavior of the process. This model is then used estimate true value of new measurements and evaluate the estimated residuals to detect and diagnose anomalies in future data (Venkatasubramanian et al., 2003c). Normally, the data-based anomaly-detection approach involves three steps: 1) model training, 2) residual generation and 3) residual evaluation. This approach uses information derived directly from input data and requires no explicit models for which development is usually costly or time consuming. Therefore, data-based methods are more attractive for practical applications to complex systems, although their performance mainly depends on the availability of an adequate amount of quality for different input data. Because each of these monitoring methods has advantages and disadvantages for different problems, the problem will determine which method is most suitable. This work focuses on the use of data-based methods for anomaly detection.

The data-based anomaly-detection techniques referenced in the bibliography can be broadly categorized into two main classes: univariate and multivariate techniques (Montgomery, 2005). Univariate statistical monitoring methods such as, the EWMA (exponentially weighted moving average) chart and CUSUM (cumulative sum) chart, are used to monitor essentially only one process variable (Harrou and Nounou, 2014). However, because modern industrial processes often present a large number of highly correlated process variables, univariate anomaly-detection methods are unable to explain different aspects of the process and, therefore, are not appropriate for modern day processes. Thus, multivariate statistical monitoring methods, which take into account correlation between process variables have developed to monitor several different process variables simultaneously. Multivariate data-based monitoring methods include latent variable methods, (e.g., partial least-squares regression (Harrou et al., 2015), principal component analysis (PCA) (Harrou et al., 2013), canonical variate analysis, independent component analysis (Chiang et al., 2001), neural networks (Neumann and Deerberg, 1999), and support vector machine based methods (Dehestani et al., 2011)). Data-based monitoring methods, especially those that utilize PCA or its extensions, have been applied across a wide range of industries, for example in the chemical industry (Simoglou et al., 1997), for water treatment (George et al., 2009), and in ecological studies (Janzekovic and Novak, 2012).

Data-based process monitoring using PCA, a well-know multivariate statistical method, has received considerable attention in the last few years (Qin, 2012; Harrou et al., 2013). PCA is a linear dimensionality reduction modeling technique that is very helpful when dealing with data sets that have a high degree of cross correlation among variables. The central idea of PCA is to reduce the dimensionality of highly correlated data, while retaining the maximum possible amount of variability present in the original data set (MacGregor and Kourti, 1995). This reduction is achieved by transforming correlated variables into a set of new uncorrelated variables, which are called principal components (PCs), each of which is a linear combination of the original variables. By reducing the dimension of the process variables, PCA is able to eliminate noise and retain only important process information; it can be employed to compress noisy and correlated measurements into a smaller informative subspace for measurement data sets. PCA-based anomaly detection is popular for use in practice because no prior knowledge about the process model is necessary and a good historical database describing the normal process operation is the only information needed.

1.2. Motivation and contribution

Detecting small or incipient anomalies in highly correlated input–output multivariate data is one of the most crucial and challenging tasks in the area of anomaly detection and diagnosis. Indeed, detection of small anomalies can provide an early warning and help to avoid catastrophic damage and subsequent financial loss. Unfortunately, conventional PCA-based monitoring indices, such as T^2 and Q statistics, often fail to detect small or moderate