ISA Transactions 41 (2002) 255–270

ISA
TRANSACTIONS®

An intelligent system for multivariate statistical process monitoring and diagnosis

Eric Tatara, Ali Çinar*

Department of Chemical and Environmental Engineering, Illinois Institute of Technology, 10 W. 33rd Street, Chicago, IL 60616, USA

(Received 26 October 2000; accepted 25 May 2001)

Abstract

A knowledge-based system (KBS) was designed for automated system identification, process monitoring, and diagnosis of sensor faults. The real-time KBS consists of a supervisory system using G2 KBS development software linked with external statistical modules for system identification and sensor fault diagnosis. The various statistical techniques were prototyped in MATLAB, converted to ANSI C code, and linked with the G2 Standard Interface. The KBS automatically performs all operations of data collection, identification, monitoring, and sensor fault diagnosis with little or no input from the user. Navigation throughout the KBS is via menu buttons on each user-accessible screen. Selected process variables are displayed on charts showing the history of the variables over a period of time. Multivariate statistical tests and contribution plots are also shown graphically. The KBS was evaluated using simulation studies with a polymerization reactor through a nonlinear dynamic model. Both normal operation conditions as well as conditions of process disturbances were observed to evaluate the KBS performance. Specific user-defined disturbances were added to the simulation, and the KBS correctly diagnosed both process and sensor faults when present. © 2002 ISA—The Instrumentation, Systems, and Automation Society.

Keywords: On-line process monitoring; Real-time monitoring; Knowledge-based systems; Multivariate statistical process monitoring; Sensor auditing; Data fusion; Canonical variate analysis; State space models

1. Introduction

Statistical methods for detecting changes in industrial processes are included in a field generally known as statistical process control (SPC) or statistical quality control. The most widely used and popular SPC techniques include univariate methods that involve observing a single variable at a given time, obtaining the mean and variance of the variable, and checking its value against upper and lower control limits. While a univariate approach may indeed work for monitoring a small number of process variables that are not correlated, current capabilities in data acquisition hardware allow a large (several thousand) number of variables to be easily measured. Application of univariate statistical process monitoring (SPM) methods to larger multivariable systems becomes difficult, if not impossible, and is often erroneous. This simplified approach to process monitoring requires an operator to continuously monitor perhaps dozens of different univariate charts, which substantially reduces the ability of plant personnel to make accurate assessments about the state of the process.

Multivariable statistical process monitoring (MSPM) techniques offer the proper theoretical framework for monitoring multivariable processes. MSPM techniques reduce the amount of raw data presented to an operator and provide a concise set of statistics that describes the process

---

*Corresponding author. Tel.: +1-312-567-3637; fax: +1-312-567-75170. *E-mail address*: cinar@iit.edu behavior. Many of the current MSPM techniques are only valid for data that are independent and identically distributed (iid). The independence assumption means that the data must not be correlated with each other, either in current measurements or lagged values, and that each have similar statistical distributions. This assumption, however, only applies to very idealized data and is not commonly observed in real processes. Correlations in the data include autocorrelation, which indicates a correlation of a single variable with its past observations, and cross-correlation, which indicates a relationship between two or more variables at either the current observations or past observation. Most process data are correlated, such as tray temperature readings in a distillation column or temperature and concentration in a chemical reactor; consequently, it is necessary to utilize MSPM methods capable of monitoring such processes. [1]

Data-driven MSPM methods that are capable of handling correlated data include methods based on principal component analysis, projection to latent structures (PLS), and canonical variate and subspace state space modeling. Data-driven techniques are dependent on data collected from a real process in order to formulate a model that describes the variability of that process. System identification methods suitable for linear processes with correlated data such as principal components regression, PLS, canonical variate state space modeling, and subspace state space modeling are described in [1–4]. System identification methodologies that can handle nonlinear systems have also been proposed [5–7]. The model developed is used to predict the future values of monitored process variables. The difference between the model predictions and the process measurement is referred to as the model residual. The residuals between the predicted and actual values may be assessed for statistical significance; a significant increase in residuals suggests the presence of abnormal conditions in the process. In addition to measuring the magnitude of the residuals, one may also examine trends in the residuals. Sensor errors in the form of a bias change, drift, or excessive noise may be diagnosed by examining the residuals over time. Periodically performing sensor audits [2,8,9] by examining the residuals for statistically significant changes can greatly enhance the credibility of the data as well as significantly reduce the number of false alarms.

The aforementioned monitoring techniques may be integrated with an automated system to provide real-time process monitoring and diagnosis. Knowledge-based systems (KBSs) or expert systems are computer systems designed in an attempt to emulate the decision-making capabilities and knowledge of a human expert in a specific field. A KBS consists of a knowledge base, decision rules, and an inferencing engine. The knowledge base is comprised of a set of knowledge, data, and facts pertaining to a specific problem and process. Decision rules are rules, developed by human experts based on intimate knowledge about the process, to reach a conclusion given the facts, such as process data. The inference engine processes the rules based on the data and conclusions reached by other rules (inferences) to reach a conclusion such as fault diagnosis.

Knowledge-based systems increased in popularity during the early 1980s when many people viewed knowledge-based systems as ‘‘magic bullet’’ solutions to every problem. Unfortunately, the high expectations of knowledge-based systems ultimately led to widespread distrust of the technology after such systems failed to perform as expected. Knowledge-based systems are gradually becoming more popular for performing such tasks as monitoring and diagnosis—tasks that KBSs are able to perform very well. Additionally, current computer hardware and software—technology that was unavailable when the KBS first gained popularity—are ideal for real-time KBS development and run-time platforms.

KBSs are useful for solving problems that can only be done by human experts or are repetitive in nature and perform effectively due to their inherently well-defined knowledge space. The advantage over human experts is access to very large databases and fast execution time. While a human expert may need to search volumes of printed text for a piece of information, a KBS can quickly search electronic databases in short time. Knowledge-based systems excel at qualitative processing, which is beneficial in diagnosis of process faults and monitoring applications. Human experts routinely perform qualitative analysis using process specific information and heuristics. Therefore, one can design a KBS to perform a specific task by transferring the knowledge from a human to a computer code. Of course, the ability of the KBS to reason is directly related to the quality of