Journal of Process Control 20 (2010) 902–911

Contents lists available at ScienceDirect

Journal of Process Control

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

Fault detection and isolation of faults in a multivariate process with Bayesian network

Sylvain Verron [a,*], Jing Li [b], Teodor Tiplica [a]

[a] LASQUO/ISTIA, University of Angers, 62, Avenue Notre Dame du Lac, 49000 Angers, France
[b] School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287-8809, United States

ARTICLE INFO

Article history:
Received 2 December 2009
Received in revised form 8 June 2010
Accepted 8 June 2010

Keywords:
Multivariate SPC
T^2 decomposition
Bayesian network

ABSTRACT

The main objective of this paper is to present a new method of detection and isolation with a Bayesian network. For that, a combination of two original works is made. The first one is the work of Li et al. [1] who proposed a causal decomposition of the T^2 statistic. The second one is a previous work on the detection of fault with Bayesian networks [2], notably on the modeling of multivariate control charts in a Bayesian network. Thus, in the context of multivariate processes, we propose an original network structure allowing to decide if a fault has appeared in the process. This structure permits the isolation of the variables implicated in the fault. A particular interest of the method is the fact that the detection and the isolation can be made with a unique tool: a Bayesian network.

1. Introduction

Nowadays, monitoring of complex manufacturing systems is becoming an essential task in order to: insure a safe production (for humans and materials), reduce the variability of products or reduce manufacturing cost. Classically, in the literature, three approaches can be found for the process monitoring [3,4]: the knowledge-based approach, the model-based approach and the data-driven approach. The knowledge-based category represents methods based on qualitative models: Digraphs; Fault Trees [5]; Case Based Reasoning [6]. The model-based approach is based on analytical (physical) models able to simulate the system [7]. Though, at each instant, the theoretical value of each sensor can be known for the normal operating state of the system. As a consequence, it is relatively easy to see if the real process values are similar to the theoretical values. However, the major drawback of this family of techniques is that a detailed model of the process is required in order to monitor it efficiently. An effective detailed model can be very difficult, time consuming and expensive to obtain, particularly for large-scale systems with many variables. The data-driven approaches are a family of different techniques based on the analysis of the real data extracted from the process [8]. These methods are based on rigorous statistical developments of the process data (i.e., control charts, methods based on Principal Component Analysis, Projection to Latent Structure or Discriminant Analysis) [3]. Since we are monitoring large multivariate processes, we will work in the data-driven monitoring framework.

To achieve this activity of data-driven monitoring, some authors call this AEM (Abnormal Event Management) [4]. This is composed of three principal steps: firstly, a timely detection of an abnormal event; secondly, diagnosing its causal origins (or root causes); and finally, taking appropriate decisions and actions to return the process in a normal working state. As the third step is specific to each process, literature generally focuses on the two first step: fault detection and diagnosis, named FDD [9]. We will call "fault" an abnormal event (like an excessive pressure in a reactor, or a low quality of a part of a product, and so on), usually defined as a departure from an acceptable range of an observed variable or a calculated parameter of the process [4]. Generally, a monitoring technique is dedicated to one specific step: detection or diagnosis. In the literature, one can find many data-driven techniques for the fault detection: univariate statistical process control (Shewhart charts) [10,11], multivariate statistical process control (T^2 and Q charts) [12,13], and some PCA (Principal Component Analysis) based techniques [14] like Multiway PCA or Moving PCA [15] used for the detection step. Kano et al. [16] make comparisons between these different techniques.

An efficient fault detection and isolation tool should be able to isolate the variables implicated in the fault, in order to help the process operator to identify the root cause (the physical cause) of the fault. Some methods exist to solve this problem (see Section 2.2), which are based on a decomposition of the T^2 statistic. But, each of these methods uses different tools for the fault detection and the fault isolation (variables implicated in the fault), like control

* Corresponding author. Tel.: +33 2 41 22 65 33; fax: +33 2 41 22 65 21.
E-mail address: sylvain.verron@univ-angers.fr (S. Verron).

0959-1524/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2010.06.001

charts, statistical decompositions, Bayesian networks, etc. From a practical point of view, it would be more interesting to combine the main advantages of these techniques and to exploit them jointly in one single tool. Recently, the application of Bayesian networks for the fault detection and diagnosis has been used with success, in the data-driven context [17,18], but also in the model-based context [19]. The objective of this article is to propose an improvement of the decomposition method of Li et al. [1], in order to use a sole Bayesian network enable to detect a fault and to isolate the implicated variables in this fault.

The article is structured as follows: Section 2 presents preliminaries needed for a correct understanding of the article; Section 2.1 highlights some aspects of Bayesian networks; Section 2.2 presents the various T^2 decompositions (causal and MYT); in Section 3 we show how to construct some multivariate control charts with a Bayesian network and how to exploit the network in order to isolate the detected faults; two examples of the approach are presented in Section 4; finally, in the last section, we conclude on the proposed approach.

2. Preliminaries

2.1. Bayesian networks

2.1.1. Definition

A Bayesian network (BN) [20,21] is an acyclic graph where each variable is a node (that can be continuous or discrete). Edges of the graph represent dependence between linked nodes. A formal definition is given here:

A Bayesian network is a triplet $\{\mathbf{G}, \mathbf{E}, \mathbf{D}\}$ where:

- $\{\mathbf{G}\}$ is a directed acyclic graph, $\mathbf{G} = (V, A)$, with V the set of nodes of $\mathbf{G}$, and A the set of edges of $\mathbf{G}$,
- $\{\mathbf{E}\}$ is a finite probabilistic space (Ω, Z, p), with Ω a non-empty space, Z a collection of subspace of Ω, and p a probability measure on Z with $p(\Omega) = 1$,
- $\{\mathbf{D}\}$ is a set of random variables associated to the nodes of $\mathbf{G}$ and defined on $\mathbf{E}$ such as:

$$p(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} p(V_i | C(V_i)) \tag{1}$$

with $C(V_i)$ the set of causes (parents) of V_i in the graph $\mathbf{G}$.

2.1.2. Dependences in Bayesian network

Theoretically, variables $X_1, X_2, \ldots, X_n$ can be discrete or continuous. But, in practice, for exact computation, only the discrete and the Gaussian case can be treated (see Ref. [22]). Such a network is often called Conditional Gaussian Network (CGN). In this context, to ensure availability of exact computation methods, discrete variables are not allowed to have continuous parents [23,24].

Practically, the conditional probability distribution is described for each node by his Conditional Probability Table (CPT) (see Ref. [22]). In a CGN, three cases of CPT can be found. The first one is for a discrete variable with discrete parents. By example, we take the case of two discrete variables A and B of respective dimensions a and b (with $a_1, a_2, \ldots, a_a$ the different modalities of A, and $b_1, b_2, \ldots, b_b$ the different modalities of B). If A is parent of B, then the CPT of B is represented in Table 1.

Table 1
CPT of a discrete node with discrete parents.

A	B			
	b_1	b_2	$\cdots$	b_b
a_1	$P(b_1\|a_1)$	$P(b_2\|a_1)$	$\cdots$	$P(b_b\|a_1)$
a_2	$P(b_1\|a_2)$	$P(b_2\|a_2)$	$\cdots$	$P(b_b\|a_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_a	$P(b_1\|a_a)$	$P(b_2\|a_a)$	$\cdots$	$P(b_b\|a_a)$

We can see that the utility of the CPT is to resume the information about the relation of B with its parent. We can denote that the dimension of this CPT is $a \times b$. In general the dimension of the CPT of a discrete node (dimension x) with p parents (discrete) $Y_1, Y_2, \ldots, Y_p$ (dimension $y_1, y_2, \ldots, y_p$) is $x \prod_{i=1}^{p} y_i$.

The second case of CPT is for a continuous variable with discrete parents. Assuming that B is a Gaussian variable, and that A is a discrete parent of B with a modalities, the CPT of B can be represented as in Table 2 where $P(B|a_1) \sim \mathcal{N}(\mu_{a_1}, \Sigma)$ indicates that B conditioned to $A = a_i$ follows a multivariate normal density function with parameters μ_{a_i} and Σ_{a_i} If we have more than one discrete parent, the CPT of B will be composed of $\prod_{i=1}^{p} y_i$ Gaussian distributions where y_i represents the respective number of modalities of the parent nodes $Y_1, Y_2, \ldots Y_p$.

The third case is when a continuous node B has a continuous parent A. In this case, we obtain a linear regression and we can write, for a fixed value a of A, that B follows a Gaussian distribution $P(B|A = a) \sim \mathcal{N}(\mu_B + \beta \times a; \Sigma_B)$ where β is the regression coefficient. Evidently, the three different cases of CPT enumerated can be combined for different cases where a continuous variable has several discrete parents and several continuous (Gaussian) parents.

The classical use of a Bayesian network (or Conditional Gaussian Network) is to enter evidence in the network (an evidence is the observation of the values of a set of variables). Thus, the information given by the evidence is propagated in the network in order to update the knowledge and obtain a posteriori probabilities on the non-observed variables. This propagation mechanism is called inference. As its name suggests, in a Bayesian network, the inference is based on the Bayes rule. Many inference algorithms (exact or approximate) have been developed, but one of the more exploited is the junction tree algorithm [25].

2.2. Existent methods for isolation

2.2.1. The MYT decomposition

As we previously stated, a method for the fault detection in multivariate processes is the T^2 control chart. However, this chart does not give any information about the diagnosis of the out-of-control situation. For that, many techniques have been proposed in the literature [26,27]. An interesting decomposition of the T^2 has been proposed by Mason et al. [26], namely MYT decomposition. The authors have proved that several isolation methods can be considered like special cases of MYT decomposition [26]. The principle of this method is to decompose the T^2 statistic in a limited number of orthogonal components which are also statistical distances. This decomposition is the following:

$$T^2 = T_1^2 + T_{2\cdot1}^2 + T_{3\cdot1,2}^2 + T_{4\cdot1,2,3}^2 + \cdots + T_{p\cdot1,2,3\cdots p-1}^2 \tag{2}$$

Table 2
CPT of a Gaussian node with discrete parents.

A	B
a_1	$P(B\|a_1) \sim \mathcal{N}(\mu_{a_1}, \Sigma_{a_1})$
a_2	$P(B\|a_2) \sim \mathcal{N}(\mu_{a_2}, \Sigma_{a_2})$
$\vdots$	$\vdots$
a_a	$P(B\|a_a) \sim \mathcal{N}(\mu_{a_a}, \Sigma_{a_a})$