We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

5,500
Open access books available

136,000
International authors and editors

170M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Fault Detection with Bayesian Network

Verron Sylvain, Tiplica Teodor, Kobi Abdessamad
LASQUO – ISTIA
France

1. Introduction

Nowadays, process control (or process monitoring) is becoming an essential task especially when dealing with complex manufacturing processes (like automatized processes containing a lot of sensors and actuators). In (Chiang et al., 2001), authors give two principal approaches to perform the process control, namely, data-driven techniques and analytical techniques. The analytical techniques are based on analytical (physical) models of the system and enable to simulate the system. 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. But, the major drawback of this family of techniques is the fact that a detailed model of the process is required in order to control 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. These methods are based on rigorous statistical development of the process data (i.e. control charts, methods based on Principal Component Analysis, Projection to Latent Structure or Discriminant Analysis) (Chiang et al., 2001).

The process control can be viewed as a three-step procedure: the fault detection, the fault diagnosis and the process recovery.

Many data-driven techniques for the fault detection can be found in the literature: univariate control charts (Shewhart charts) (Shewhart, 1931), multivariate control charts (T^2, Q, MEWMA, MCUSUM charts) (Hotelling, 1947; Lowry et al., 1992; Pignatiello & Runger, 1990), and some PCA (Principal Component Analysis) based techniques (Jackson, 1985), like Moving PCA (Bakshi, 1998). In (Kano et al., 2002), authors make comparisons between these different techniques. Other important approaches are PLS (Projection to Latent Structures) based approaches (MacGregor & Kourti, 1995).

In order to accomplish the fault diagnosis, many approaches have been proposed (Kourti & MacGregor, 1995). The fault diagnosis procedure can also be considered as a classification task. Many classifiers have been developed (i.e. Fisher Discriminant Analysis (Duda et al., 2001), Support Vector Machine (Vapnik, 1995), k-nearest neighborhood (Cover & Hart, 1967), Artificial Neural Networks (Bishop, 2001) and bayesian classifiers (Friedman et al., 1997)).

When dealing with a great number of variables and faults, the performances of these classifiers are diminished. To overcome the inconvenience, a feature selection is often performed before the classification task. By feature selection one means the identification of the key process variables allowing capturing the signature (identity) of the various faults affecting the process.

In this article, we are presenting a new method for the detection and the diagnosis of faults in multivariate processes with bayesian networks.

The chapter is structured as follows: the second section present the utilization of the multivariate control charts for fault detection; the third section highlights some aspects on bayesian networks and particularly on bayesian network classifiers; in the fourth section we are showing how to model some multivariate control charts (T^2 and MEWMA), with a bayesian network. In the last section, we conclude on the proposed approach and give some outlooks.

2. Detection and classification

As the new method for the fault detection that we proposed is based on multivariate control charts and on discriminant analysis, we present this two techniques in this section.

2.1 Multivariate control charts

Hotelling (Hotelling, 1947) was the first to propose a control chart for the monitoring of multivariate processes: the T^2 control chart. (Hawkins, 1991) demonstrated that this chart represents the best statistical test for the detection of a change in the mean of a multivariate process. For a process with p variables, we can write the T^2 statistics as the following:

$$T^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \tag{1}$$

where: $\mathbf{x}$ is the observation vector of size $1 \times p$, $\boldsymbol{\mu}$ is the mean vector of size $1 \times p$, $\boldsymbol{\Sigma}$ is the variance-covariance matrix of size $p \times p$ and the symbol ^{T} represents the transpose of a vector or a matrix.

For each instant sampling, the T^2 will be compared to an upper control limit CL (lower control limit is fixed to 0) in order to conclude about the state of the process: if the value of T^2 is lower than the CL, then the process is "in control", either ($T^2 > CL$) the process is declared out of control signifying a fault has occurred in the process. The computation of the limit will depend of the estimation of the process parameters. Montgomery (Montgomery, 1997) gives the different computations for this limit.

The major drawback of the T^2 control chart is its moderate performances to detect small mean shifts. In order to solve this problem, other multivariate control charts have been proposed: MEWMA (Multivariate Exponentially Weighted Moving Average) (Lowry et al., 1992) and MCUSUM (Multivariate CUmulative SUM) (Pignatiello & Runger, 1990). These charts are respectively the multivariate analogous of the EWMA and CUSUM control charts. The principle of the MEWMA control chart is to take into account the process evolution in weighting past observations extracted from the process. So, the MEWMA variable $\mathbf{y}_t$ is

computed recursively, for each sample, by the equation (2) where the initialization is given by $\mathbf{y}_0 = \boldsymbol{\mu}$.

$$\mathbf{y}_t = \boldsymbol{\Lambda}\mathbf{x}_t + (\mathbf{I} - \boldsymbol{\Lambda})\mathbf{y}_{t-1} \qquad (2)$$

In equation (2), $\mathbf{x}_t$ is the observation vector at instant t, $\mathbf{I}$ is the identity matrix and $\boldsymbol{\Lambda}$ the weighting diagonal matrix which elements $\lambda_1, \lambda_2, \cdots, \lambda_p$ are comprised between 0 and 1. ($0 < \lambda_i \leq 1$).

Lowry proposed to use $\lambda_i = \lambda$ for $i = 1, 2, \cdots, p$ if there is no particular reason to differently weight the variables. In this case, the MEWMA control chart is considered "directionally invariant" because no direction in the multivariate space is advantaged. By choosing different λ_i in order to increase the detection performance in certain directions, one obtains a "directional" control chart which is quite more difficult to compute and to interpret. So, in the rest of this chapter, we will consider the non-directional MEWMA control chart ($\lambda_i = \lambda$ for all i). Thus, we can rewrite equation (2) as:

$$\mathbf{y}_t = \lambda\mathbf{x}_t + (1 - \lambda)\mathbf{y}_{t-1} \qquad (3)$$

On the MEWMA control chart, we plot the following statistic:

$$T_t^2 = (\mathbf{y}_t)^T \boldsymbol{\Sigma}_{\mathbf{y}_t}^{-1} (\mathbf{y}_t) \qquad (4)$$

Where $\boldsymbol{\Sigma}_{\mathbf{y}_t}$ is the variance-covariance matrix of the variable $\mathbf{y}$ at instant t. This matrix is defined as:

$$\boldsymbol{\Sigma}_{\mathbf{y}_t} = \left\{ \frac{\lambda\left[1 - (1 - \lambda)^{2t}\right]}{2 - \lambda} \right\} \boldsymbol{\Sigma} \qquad (5)$$

But, if λ is not too small ($\lambda > 0.1$), this matrix approaches rapidly (in 4 or 5 samples) his asymptotic value defined by:

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \left\{ \frac{\lambda}{2 - \lambda} \right\} \boldsymbol{\Sigma} \qquad (6)$$

The process is declared out-of-control if T_t^2 is greater than a control limit h_M. This limit is function of p and λ, in order to respect a given false alarm frequency (Lowry et al., 1992). We can precise that performances of the MEWMA control chart are function of λ. Indeed, a small λ allows a performing detection of small magnitude shifts, but a higher λ will be more adapted for large magnitude shifts. So, the choice of λ will be function of the

magnitude shift that one wants to detect. A particular case of the MEWMA control chart is the case where $\lambda=1$. In this case, the MEWMA chart is equivalent to the T^2 control chart.

2.2 Discriminant analysis

The discriminant analysis can be viewed as a statistical technique for supervised classification, based on the Bayes rule. Indeed, for a problem with k classes, this technique allocates to a new observation $\mathbf{x}$ the class C_i having the maximal a posteriori probability ($P(C_i|\mathbf{x})$):

$$\mathbf{x} \in C_i \text{, if } i=\underset{i=1,\cdots,k}{\operatorname{argmax}}\left\{P(C_i|\mathbf{x})\right\} \tag{7}$$

This rule is called "Bayes rule" because it allows obtaining the value of $P(C_i|\mathbf{x})$ with the use of equation (8), where $P(C_i)$ represents the a priori probability of the class C_i.

$$P(C_i|\mathbf{x})=\frac{P(C_i)P(\mathbf{x}|C_i)}{P(\mathbf{x})} \tag{8}$$

We can see that for each class, the denominator of equation (8) is the same and it will not interfere in the discriminant function. So, the equation (7) can be rewritten as:

$$\mathbf{x} \in C_i \text{, if } i=\underset{i=1,\cdots,k}{\operatorname{argmax}}\left\{P(C_i)P(\mathbf{x}|C_i)\right\} \tag{9}$$

But, for more simplicity, we will write this decision rule under the form of a cost function K as given by:

$$K_i(\mathbf{x})=-2\log(P(C_i)P(\mathbf{x}|C_i)) \tag{10}$$

Thus, the attribution rule of a new observation $\mathbf{x}$ to a class C_i can be written by the following equation:

$$\mathbf{x} \in C_i \text{, if } i=\arg\min_{i=1,\cdots,k}\left\{K_i(\mathbf{x})\right\} \tag{11}$$

Generally, this rule is applied with parametric distribution laws; the most used is the multivariate normal (Gaussian) distribution.
The multivariate Gaussian distribution is the generalization of the one-dimensional normal distribution to higher dimensions. The probability density function ϕ, conditionally to a

class C_i is given in equation (12), where $\boldsymbol{\mu}_i$ is the mean vector of the class C_i, and $\boldsymbol{\Sigma}_i$ is the variance-covariance matrix of the class C_i.

$$\phi(\mathbf{x}|C_i) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right) \tag{12}$$

As the exact values of $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ are generally unknown, it is necessary to estimate them with data. In order to estimate these parameters, one can use MLE (Maximum Likelihood Estimation). An advantage of MLE is the fact that this estimator has good convergence propriety when the sample size is large enough. More, this is one of the simplest techniques for the estimation of law parameters. In the case of the multivariate Gaussian distribution, the MLE of the mean vector is:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \tag{13}$$

In the same way, an unbiased estimation of $\boldsymbol{\Sigma}$ is:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\hat{\boldsymbol{\mu}})(\mathbf{x}_i-\hat{\boldsymbol{\mu}})^T \tag{14}$$

For a more detailed justification of these estimators, see (Duda et al., 2001).
The discriminant analysis rule (equation 10) applied in the case of the multivariate Gaussian distribution can be developed as:

$$K_i(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - 2\log(P(C_i)) + \log(|\boldsymbol{\Sigma}_i|) + p\log(2\pi) \tag{15}$$

In the equation, we can see that the last term ($p\log(2\pi)$) is constant for each K_i and we have not to take it into account for the discrimination. This rule is named "Quadratic Discriminant Analysis". We can also remark that $(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$ is the T^2 of $\mathbf{x}$ for the class C_i. If the assumption of independent variables is made, thus $\boldsymbol{\Sigma}_i$ is a diagonal matrix (all the covariances are null). This decision rule is also known as the Bayes classifier or naïve bayesian network. The major problem of the quadratic discriminant analysis is that this technique requires an important amount of data for the correct estimation of all the parameters. In the case of non-sufficient data, we can use the linear discriminant analysis.
The important assumption of the linear discriminant analysis, compared to the quadratic one, is the supposed equality of all variance-covariance matrices. So, for each class C_i, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, with $\boldsymbol{\Sigma}$ the pooled sample covariance matrix. This matrix is easily obtained with

equation (16) where n_i is the sample number of the class C_i and where n is the total number sample (so $n = n_1 + n_2 + \ldots + n_k$).

$$\Sigma = \frac{(n_1-1)\Sigma_1 + (n_2-1)\Sigma_2 + \ldots + (n_k-1)\Sigma_k}{n-k} \tag{16}$$

So, in this case, equation (15) becomes:

$$K_i(\mathbf{x}) = (\mathbf{x}-\mathbf{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x}-\mathbf{\mu}_i) - 2\log(P(C_i)) + cst \tag{17}$$

where $\log(|\Sigma|) + p\log(2\pi) = cst$. One can see in equation (17) that if the a priori probabilities ($P(C_i)$) of each class are equal, thus the decision rule comes to the computation of the T^2 for each class and the attribution of a new observation to the class with the lower T^2. This cost function makes linear bound between the different classes. But, as for the case of quadratic discriminant analysis, we can do the assumption that Σ is diagonal. The linear discriminant analysis is quite robust to the assumption of normality of each class and to assumption of equality of the different variance-covariance matrices. For these reasons, this technique is widely used and is considered as a reference method of supervised classification.

3. Discriminant analysis and bayesian networks

An interesting tool using statistics is bayesian network, an oriented probabilistic graphic model. Bayesian networks can be efficient supervised classifiers. So, after the general presentation of bayesian networks and especially bayesian classifiers, we will present how to make discriminant analysis with them.

3.1 Bayesian networks

A bayesian network (Pearl, 1988; Jensen, 1996) is a triplet {**G, E, D**} where:
{**G**} is a directed acyclic graph, **G=(V,A)**, where **V** is the set of nodes of **G**, and **A** is the set of edges of **G**,
{**E**} is a finite probabilistic space (Ω, Z, P), where Ω is a non-empty space, Z is a collection of subspace of Ω, and P is a probability measure on Z with $P(\Omega)=1$,
{**D**} is a set of random variables associated to the nodes of **G** and defined on **E** such as:

$$P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P(V_i | C(V_i)) \tag{18}$$

where $C(V_i)$ is the set of parents of V_i in the graph **G**.
Bayesian network classifiers are particular bayesian networks (Friedman et al., 1997). They always have a discrete node C coding the k different classes of the system. The remaining variables X_i represent the descriptors (variables) of the system.

A Naïve Bayesian Network (NBN) is a particular type of bayesian network classifiers (Langley et al., 1992). It is also known as the Bayes classifier. In a NBN, the class node is linked with all other variables of the system (descriptors) as indicated on the figure 1.
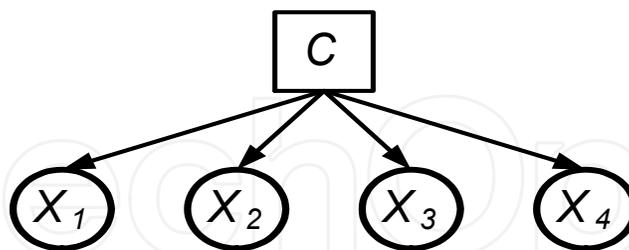

Fig. 1. Example of a Naïve Bayesian Network (NBN)

The NBN is called naïve because it makes the naïve (but strong) assumption that all descriptors (variables of the system) are class conditional statistically independent (no correlation between each descriptor in each class). In (Inza et al., 1999), authors made a comparative study of classification algorithms used in artificial intelligence. NBN is compared with other methods like k-nearest neighborhood, C4.5, decision tree and so on. If the independence assumption of the descriptors is verified and that probabilities are estimated with enough precision the NBN is an optimal classifier in term of misclassification rate (Domingos & Pazzani, 1996). This optimality is obtained when continuous variables are discretized in such a way that every distribution can be well approximated by discretization. Of course, the discretization of variables is a loss of information because it is a reduction of the variables space. But, assuming that continuous variables follow normal probability density function, we can deal with them directly in the network. And, if this assumption is verified, we keep this optimality. But, in many systems, it is very frequent to have high correlations between variables, and a NBN will not take into account these correlations. Extensions of NBN have been developed in order to solve this problem.

A first interesting extension is the TAN (Tree-Augmented bayesian Network) (Friedman et al., 1997). In a TAN, a maximum weighted spanning tree is constructed with the descriptors following the algorithm of (Chow & Liu, 1968). So, each descriptor will have at most one other descriptor as parent. After that, edges from the class node to each descriptor are added (like a NBN). An example of a TAN is given on the figure 2.
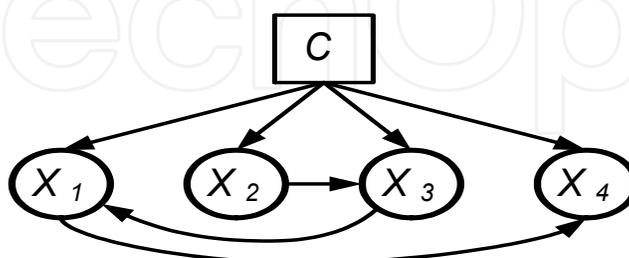

Fig. 2. Example of a Tree-Augmented bayesian Network (TAN)

Another extension is the k-dependence bayesian classifier structure (kDB structure) (Sahami, 1996) that extends the TAN structure allowing a maximum of k predictor parents plus the class for each predictor variable (TAN structures are equivalent to kDB structures

with k = 1). Finally, the bayesian multinets (Geiger & Heckerman, 1996) introduce a different structure for each value of the class variable (a particular case is to take a different TAN for each value of the class). In (Friedman et al., 1997), authors show that these classifiers give a lower misclassification rate than the NBN. But, these classifiers do not take into account all the correlations between variables.

An other extension to the NBN is the Condensed Semi Naïve Bayesian Network (CSNBN) (Kononenko, 1991). The principle of this classifier is to represent some variables in a joint node (i.e. some normally distributed variables can be modeled with a node representing a multivariate normal distribution). In this way, all correlations of the system will be taken into account. A CSNBN will be composed of two nodes: the class node and a multivariate node. An example of a CSNBN is given on the figure 3.
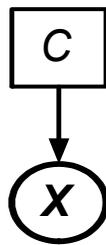
Fig. 3. Example of a Condensed Semi Naïve Bayesian Network (CSNBN)

3.2 Discriminant analysis as a bayesian network

Discriminant analysis techniques can be easily transposed to a bayesian network. Indeed, inference in a bayesian network is based on the Bayes rule, like the discriminant analysis (see section 2.2). The structure of the network in order to make classical discriminant analysis (quadratic or linear) on a system with p variables and k classes can be modelized with two linked nodes. The first one (node C) is a discrete node (with k modalities) representing the classes of the system. The second one (node $\mathbf{X}$) is a multivariate Gaussian node. We have previously presented this type of bayesian network: it is a CSNBN (figure 4). This network represents a multivariate normal law conditionally to the class, as the discriminant analysis. The choice between the different types of discriminant analysis (quadratic, linear, spherical, diagonal, etc) is made on the choice of the k variance-covariance matrices that will be attributed to $\mathbf{X}$. It is also possible to make a quadratic discriminant analysis in viewing all the univariate variables X_i composing the multivariate variable $\mathbf{X}$, and then, linking all them in order to take into account all possible relations between these univariate variables X_i. A four variable example of such structure is given on figure 4.
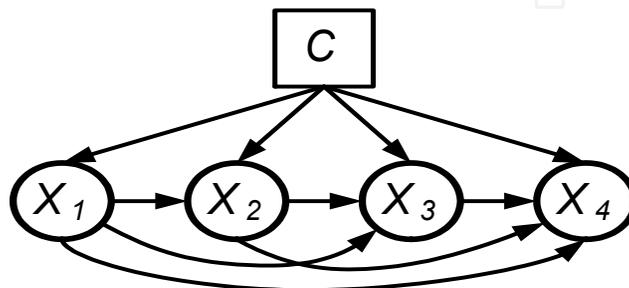
Fig. 4. Bayesian Network for Discriminant Analysis with incomplete observations

This figure 4 represents a succession of *p-1* linear regressions. The interest of this type of structure is the fact that it can give response even if the values of some X_i are unknown. However, the estimation of all the regressions parameters and the inference time will be more consequent than for the CSNBN. So, in the basic case (complete data), we will prefer CSNBN.

4. Multivariate control charts with a bayesian network

The detection, as we previously said, consists in identifying the presence of faults in the process. The detection can be considered as a classification in two classes: no fault in the process, and fault in the process. For the detection, we suppose to have samples for the normal working state of the system (we will call this class "In Control" (IC) class), but having fault samples should not be necessary. This classification type is called one-class classification (Tax & Duin, 2001). In contrast with normal classification problems where one tries to distinguish between two (or more) classes of objects, one-class classification tries to describe one class of objects, and distinguish it from all other possible objects. A useful solution for this is to create at least a second class, a virtual class named "Out-of-Control" (OC) class. This virtual class represents the set of observation that cannot be attribute to the IC class.

An example of one-class classification is the T^2 control chart (new observation is attributed to the IC class if his T^2 is lower than CL, and to the OC class if his T^2 is greater than CL). In this case, it is evident that the decision boundary can be represented as the CL. For this control chart, it is possible to represent, for a bidimensional example, the decision boundary induced by CL. On the figure 5, we can see that this boundary is an ellipse rounding the normal working class of the process (IC class), and that each observation out of this ellipse is attributed to the fault class (OC class).
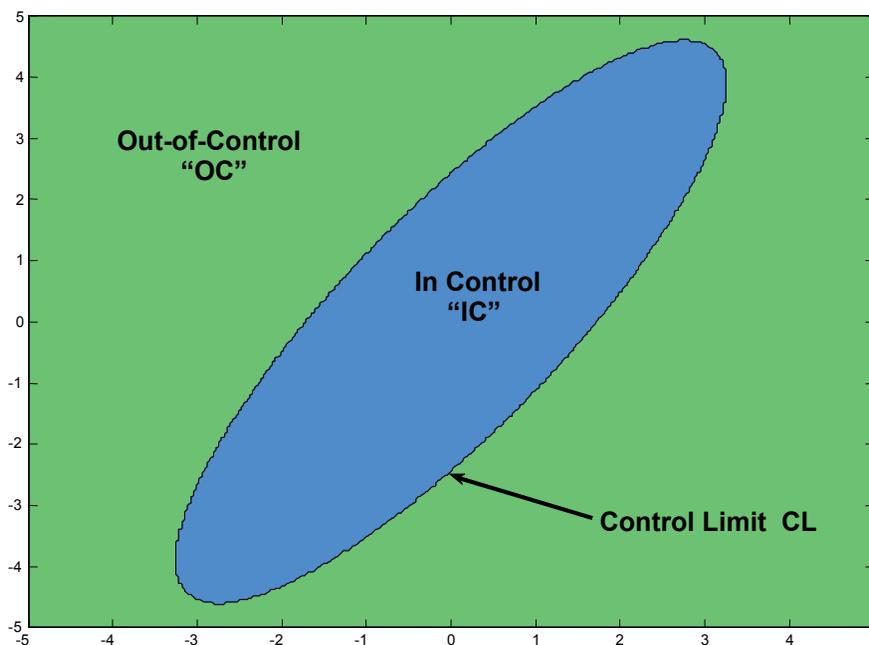

Fig. 5. Decision boundary of the control chart for a bivariate system

On the figure 5, we can see that the classification technique producing this boundary cannot be linear. So, it will be impossible to make such boundary with linear discriminant analysis. But, this boundary shape is typical of quadratic discriminant analysis.

In order to define the In Control class, we make the assumption that some samples of in control data are available. In the same way that for the control chart, these samples allow to estimate the mean vector μ and the variance-covariance matrix Σ of the process in normal working conditions. The major problem is the definition of the virtual class OC because no sample is available. In studying the figure 5, we can see that if we make the assumption that the mean of the OC class is the same that the mean of the IC class, the only difference between the two classes is the fact that the OC class has more variability that the IC class. So, we will define the virtual class as following: the mean vector of the OC class is equal to the mean vector of the IC class; the variance-covariance matrix of the OC class is equal to $c\Sigma$ where Σ is the variance-covariance matrix of the IC class, and where c is a coefficient strictly greater than 1. This coefficient c allows increasing the variability of the OC class in function of the variability of the IC class. The table 1 presents the parameters of each class.

Class	Distribution
In Control (IC)	$\mathbf{X} \sim N\ (\mu, \Sigma)$
Out-of-Control (OC)	$\mathbf{X} \sim N\ (\mu, c \times \Sigma)$

Table 1. Classes parameters

However, fixing the different distributions laws is not sufficient for taking decisions about the process state (in or out-of-control). Indeed, the definition of these parameters makes possible the application of a discriminant analysis in a bayesian network. So, when a new observation $\mathbf{x}$ is injected in the bayesian network, we will obtain an a posteriori probability $P(IC|\mathbf{x})$ that this observation belongs to the IC class and an a posteriori probability $P(OC|\mathbf{x})$ that this observation belongs to the OC class (with $P(IC|\mathbf{x}) + P(OC|\mathbf{x}) = 1$). In order to decide about the state of the process, one has to fix a false alarm rate α in addition to these two previous calculated probabilities. As in the case of multivariate control charts, we have to fix a false alarm rate. Moreover, the coefficient c will play an important role in the probabilities computations. In the next section, we will see how to set the different parameters to obtain the equivalency between the bayesian network and the multivariate control charts.

4.1 Equivalency proof

As in the case of the multivariate control charts (T^2 or MEWMA), we will fix a threshold (limit) on the a posteriori probabilities allowing to take decisions on the process: if, for a given observation, the a posteriori probability to be out-of-control ($P(OC|\mathbf{x})$) is greater than the a priori probability to be out-of-control ($P(OC)$), then this observation is out-of-control. This rule can be rewritten as: "process out-of-control if $P(OC|\mathbf{x}) > P(OC)$", or equivalently "process in control if $P(IC|\mathbf{x}) > P(IC)$". The objective of the following developments is to define c in order to obtain the equivalency between the bayesian network and the multivariate control charts.

We want to keep the following decision rule:

$$\mathbf{x} \in \text{IC, if } T^2 < CL \tag{19}$$

with this decision rule:

$$\mathbf{x} \in \text{IC, if } P(IC|\mathbf{x}) > P(IC) \tag{20}$$

We develop the second decision rule:

$$
\begin{aligned}
P(IC|\mathbf{x}) &> P(IC) \\
P(IC|\mathbf{x}) &> (P(IC))(P(IC|\mathbf{x}) + P(OC|\mathbf{x})) \\
P(IC|\mathbf{x}) &> (P(IC))P(IC|\mathbf{x}) + (P(IC))P(OC|\mathbf{x}) \\
P(IC|\mathbf{x}) &> \left(\frac{P(IC)}{1 - P(IC)}\right)P(OC|\mathbf{x}) \\
P(IC|\mathbf{x}) &> \left(\frac{P(IC)}{P(OC)}\right)P(OC|\mathbf{x})
\end{aligned}
\tag{21}
$$

But, the Bayes law gives:

$$P(IC|\mathbf{x}) = \frac{P(IC)P(\mathbf{x}|IC)}{P(\mathbf{x})} \tag{22}$$

And

$$P(OC|\mathbf{x}) = \frac{P(OC)P(\mathbf{x}|OC)}{P(\mathbf{x})} \tag{23}$$

So, we obtain:

$$
\begin{aligned}
\frac{P(IC)P(\mathbf{x}|IC)}{P(\mathbf{x})} &> \left(\frac{P(IC)}{P(OC)}\right)\frac{P(OC)P(\mathbf{x}|OC)}{P(\mathbf{x})} \\
\left(\frac{P(IC)}{P(OC)}\right)P(\mathbf{x}|IC) &> \left(\frac{P(IC)}{P(OC)}\right)P(\mathbf{x}|OC) \\
P(\mathbf{x}|IC) &> P(\mathbf{x}|OC)
\end{aligned}
\tag{24}
$$

In the case of a discriminant analysis with k classes C_i, the conditional probabilities are computed with equation (25), where ϕ represents the probability density function of the multivariate Gaussian distribution of the class.

$$P(\mathbf{x}|C_i)=\frac{\phi(\mathbf{x}|C_i)}{\sum\limits_{j=1}^{k} P(C_j)\phi(\mathbf{x}|C_j)} \tag{25}$$

So, equation (24) can be written as:

$$\phi(\mathbf{x}|IC)>\phi(\mathbf{x}|OC) \tag{26}$$

We recall that the probability density function of a multivariate Gaussian distribution of dimension p, of parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, of an observation $\mathbf{x}$ is given by:

$$\phi(\mathbf{x})=\frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \tag{27}$$

If the law parameters are $\boldsymbol{\mu}$ and $c\times\boldsymbol{\Sigma}$, then the density function becomes:

$$\phi(\mathbf{x})=\frac{\exp\left(-\frac{1}{2c}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}c^{p/2}} \tag{28}$$

In identifying the expression $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ as the T^2 of the observation $\mathbf{x}$, we can write:

$$\phi(\mathbf{x}|IC)>\phi(\mathbf{x}|OC)$$

$$\frac{\exp\left(-\frac{T^2}{2}\right)}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}>\frac{\exp\left(-\frac{T^2}{2c}\right)}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}c^{p/2}}$$

$$\exp\left(-\frac{T^2}{2}\right)>\frac{\exp\left(-\frac{T^2}{2c}\right)}{c^{p/2}} \tag{29}$$

$$\frac{T^2}{2}>\frac{T^2}{2c}-\frac{p\ln(c)}{2}$$

$$T^2<\frac{p\ln(c)}{1-\frac{1}{c}}$$

However, we search the value(s) of c allowing the equivalency with the control chart decision rule: $\mathbf{x} \in$ IC, if $T^2<CL$. So, we obtain the following equation for c:

$$\frac{p\ln(c)}{1-\frac{1}{c}}=CL \qquad (30)$$

Or, equivalently:

$$1-c+\frac{pc}{CL}\ln(c)=0 \qquad (31)$$

Equation (31) admits two solutions: $c=1$ (not acceptable) and a second solution (numerically computable) which depends of p and α. With the coefficient c correctly computed, we obtain the equivalence between the bayesian network and the multivariate control charts. We precise that, as univariate charts are simply a particular case of multivariate control chart, the proof given is also available for univariate control charts. In order to demonstrate the proposed approach, we illustrate it on a simple system with two variables.

4.2 Detection with bayesian network

We will study a T^2 control chart and a MEWMA control chart (with $\lambda=0.1$) modelized by bayesian networks. We choose a false alarm rate $\alpha=1\%$. When the system is in-control, it follows a multivariate Gaussian distribution with parameters $\mathbf{\mu}$ and $\mathbf{\Sigma}$ such as:

$$\mathbf{\mu}=(5\ 10) \qquad (32)$$

$$\mathbf{\Sigma}=\begin{pmatrix} 1 & 1.2 \\ 1.2 & 2 \end{pmatrix} \qquad (33)$$

In order to monitor this process, we apply the proposed method of detection with bayesian network. So, for a T^2 control chart, we obtain the bayesian network of the figure 6. We have also given the conditional probability table of each node, and where c is equal to 95.28 (solution of equation (31) for $\alpha=1\%$ and $p=2$).
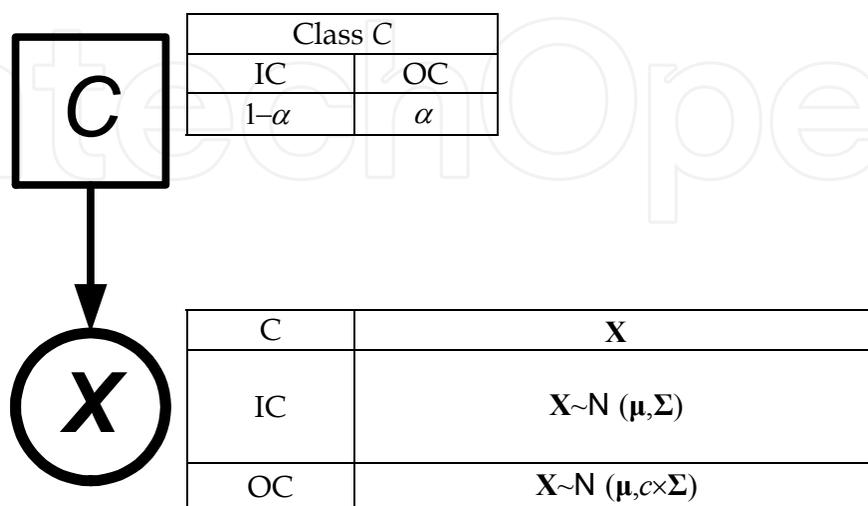

Class C	
IC	OC
$1-\alpha$	α

C	X
IC	$\mathbf{X} \sim N\,(\mathbf{\mu},\mathbf{\Sigma})$
OC	$\mathbf{X} \sim N\,(\mathbf{\mu},c\times\mathbf{\Sigma})$

Fig. 6. Bayesian Network similar to T^2 control chart

In the same way, we can also monitor the process with a MEWMA control chart modelized by the bayesian network of the figure 7, where c is equal to 90.29 (solution of equation (31) for α=1% and p=2 in the MEWMA case).
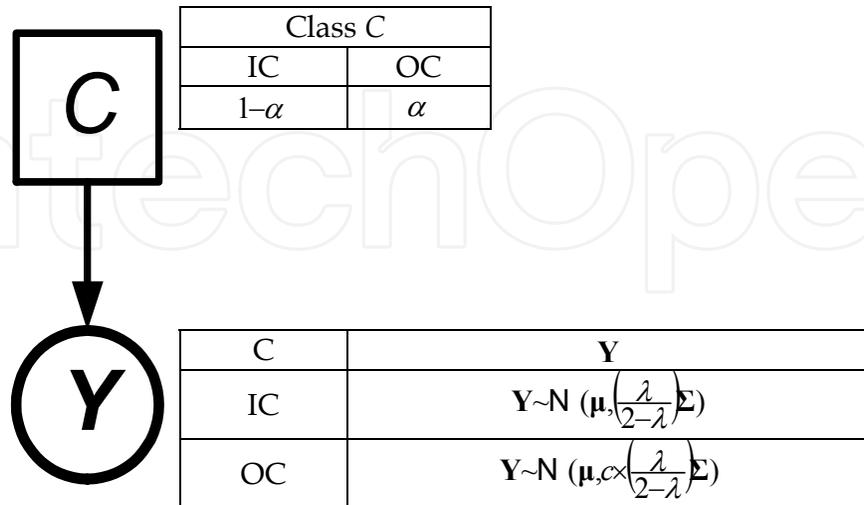
Class C	
IC	OC
$1-\alpha$	α

C	Y
IC	$\mathbf{Y} \sim N\ (\boldsymbol{\mu}, \left(\dfrac{\lambda}{2-\lambda}\right)\boldsymbol{\Sigma})$
OC	$\mathbf{Y} \sim N\ (\boldsymbol{\mu}, c \times \left(\dfrac{\lambda}{2-\lambda}\right)\boldsymbol{\Sigma})$

Fig. 7. Bayesian Network similar to MEWMA control chart

We have simulated this system on 30 observations. But, a fault has been introduced from observation 6 to 30. This fault is a mean step of magnitude 0.5 on the first variable. The figure 8 represents the decision taken at each instant respectively for the T^2 chart (left graphs) and for the MEWMA chart (right graphs). On this figure, upper graphs represent the computation of the statistical distance associated with the control chart (T^2 or T_t^2). The lower graphs give the a posteriori probability to be in control. The control limit is given on each graph, so we can view the limit on the bayesian network fixed to $1-\alpha$=99% .
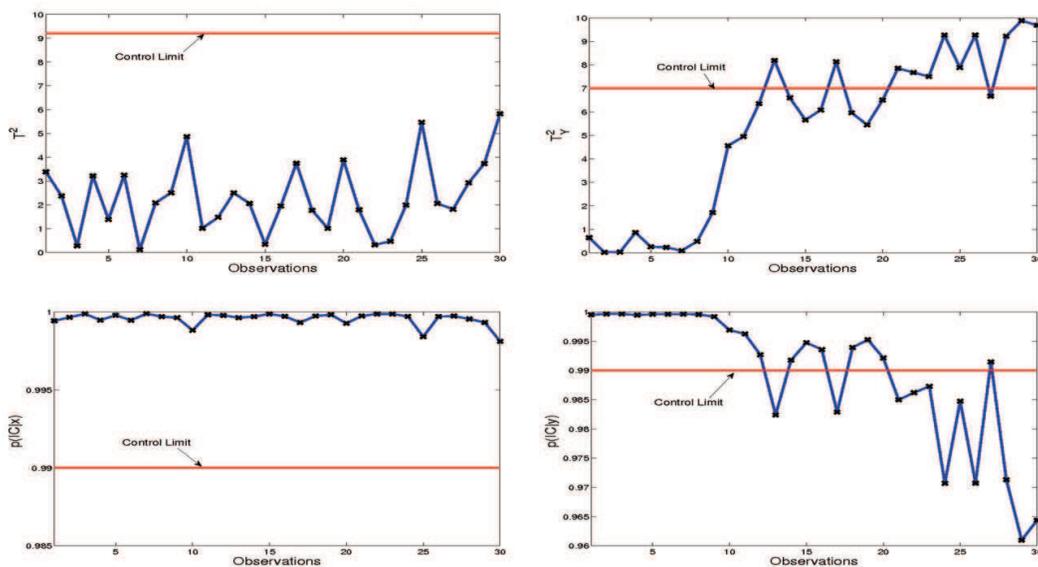
Fig. 8. Results of the T^2 and MEWMA chars, and their equivalency in Bayesian Network

On the figure 8, we can see that for each instant, the decision between a control chart and its modelization by bayesian network is equivalent.

We demonstrated that it is possible to have detection of faults in multivariate processes with bayesian networks and we proved that we can easily modelize multivariate control charts with them.

5. Conclusions and outlooks

In this chapter, we show that a bayesian network can be an efficient way to diagnose a fault in multivariate processes. We have selected two statistical fault detection techniques (the T^2 chart and the MEWMA chart) and we have demonstrated that these charts can be viewed as a discriminant analysis and so can be implemented in a simple bayesian network.

As the efficiency of bayesian network for the diagnosis of systems has already been demonstrated (Verron et al., 2006; Verron et al., 2007), the evident outlook of this work is the full study of the use of bayesian network in order to monitor and control a multivariate process (detection and diagnosis in the same network).

6. References

Bakshi, B. R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, Vol. 44, No. 7, pp. 1596-1610.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, Vol. 13, pp. 21-27.

Chiang, L.H.; Russell, E.L. & Braatz, R.D. (2001). *Fault detection and diagnosis in industrial systems*, Springer-Verlag, New York.

Chow, C. & Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, Vol. 14, pp. 462-467.

Domingos, P. & Pazzani, M.J. (1996). Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier, *In Proceedings of the Thirteen International Conference on Machine Learning*, pp. 105-112.

Duda, R.O.; Hart, P.E. & Stork, D.G. (2001). *Pattern Classification 2nd edition*, Wiley.

Friedman, N.; Geiger, D.& Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, Vol. 29, No. 2, pp 131-163.

Geiger, D. & Heckerman D. (1996). Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, Vol. 82, pp. 45-74.

Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics*, Vol. 33, pp. 61-75.

Hotelling, H. (1947). Multivariate Quality Control. In *Techniques of Statistical Analysis*, C. Eisenhart; M.W. Hastay & W.A. Wallis, pp. 111-184, McGraw-Hill, New York.

Inza, I.; Larranaga, P.; Sierra, B.; Etxeberria, R.; Lozano, J. & Pena, J. (1999). Representing the behaviour of supervised classification learning algorithms by Bayesian networks. *Pattern Recognition Letters*, Vol. 20, pp. 1201-1209.

Jackson, E.J. (1985). Multivariate quality control. *Communication Statistics - Theory and Methods*, Vol. 14, No. 2, pp. 657-688.

Jensen, F.V. (1996). *An introduction to Bayesian Networks*, Taylor and Francis, London.

Kano, M.; Nagao, K.; Hasebe, S.; Hashimoto, I.; Ohno, H.; Strauss, R.; & Bakshi, B. (2002). Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. *Computers and Chemical Engineering*, Vol. 26, No. 2, pp. 161-174.

Kononenko, I. (1991). Semi-naive bayesian classifier, *In Proceedings of the Sixth European Working Session on Learning*, pp. 206-219, Porto, Portugal, Springer-Verlag.

Kourti, T. & MacGregor, J.F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, Vol. 28, No. 4, pp. 409-428.

Langley, P; Iba, W. & Thompson, K. (1992). An analysis of bayesian classifiers, *In Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 223-228, San Jose, CA: AAAI Press.

Lowry, C.A.; Woodall, W.H.; Champ, C.W. & Rigdon, S.E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, Vol. 34, No. 1, pp46-53.

MacGregor, J. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, Vol. 3, No. 3, pp. 403-414.

Montgomery, D.C. (1997). *Introduction to Statistical Quality Control, Third Edition*, John Wiley and Sons.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers.

Pignatiello, J. & Runger, G. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, Vol. 22, No. 3, pp. 173-186.

Sahami, M. (1996). Learning Limited Dependence Bayesian Classifiers, *In Proceedings of the Second International Conference on Knowledge Discovery in Databases*, pp. 335-338.

Shewhart, W.A. (1931). *Economic control of quality of manufactured product*, D. Van Nostrand Co., New York.

Tax, D. M. J. & Duin, R. P. W. (2001). Combining One-Class Classifiers. *Lecture Notes in Computer Science*, Vol. 2096, pp. 299-308.

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory, Springer.

Verron S., Tiplica T., Kobi A. (2006). A new procedure based on mutual information for fault diagnosis of industrial systems, *in Workshop on Advanced Control and Diagnosis*.

Verron S., Tiplica T., Kobi A. (2007). Fault diagnosis of industrial systems with bayesian networks and mutual information, *In Proceedings of the Ninth International European Control Conference*.
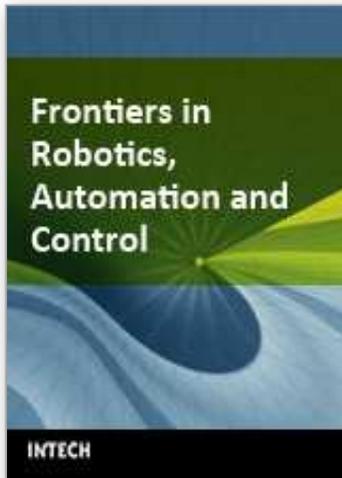
Frontiers in Robotics, Automation and Control

Edited by Alexander Zemliak

This book includes 23 chapters introducing basic research, advanced developments and applications. The book covers topics such us modeling and practical realization of robotic control for different applications, researching of the problems of stability and robustness, automation in algorithm and program developments with application in speech signal processing and linguistic research, system's applied control, computations, and control theory application in mechanics and electronics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Verron Sylvain, Tiplica Teodor and Kobi Abdessamad (2008). Fault Detection with Bayesian Network, Frontiers in Robotics, Automation and Control, Alexander Zemliak (Ed.), ISBN: 978-953-7619-17-6, InTech, Available from:
http://www.intechopen.com/books/frontiers_in_robotics_automation_and_control/fault_detection_with_bayesian_network

INTECH
open science | open minds