FAULT DETECTION IN A MULTIVARIATE PROCESS WITH A BAYESIAN NETWORK

VERRON Sylvain – TIPLICA Teodor – KOBI Abdessamad

LASQUO / ISTIA
University of Angers
62, Avenue Notre Dame du Lac
49000 – Angers – France
sylvain.verron@istia.univ-angers.fr

Abstract: The purpose of this article is to present a new method for the fault detection of a multivariate process. This method is based on the utilization of a bayesian network in order to monitor the process evolution. The class node of the network corresponds to the state of the process (in control or out of control) in probability term. The other nodes correspond to the process values for different instants. A threshold is fixed with simulation so as to respect a given average run length. This threshold permits to conclude of the process state. The method is evaluated with simulations in order to analyze and compare his performances to other multivariate chart T^2 of Hotelling and MEWMA.

Keywords: SPC, multivariate, detection, bayesian network, T^2, MEWMA

1 – Introduction

Control charts have been widely used in industry. The aim is to monitor the centering and the scattering of the process. It was in 1931 that Shewart [Shewhart] has suggested the first control chart $\overline{X}$. In spite of years and the development in the informatics field, this control chart is used yet nowadays. But, the Shewhart chart and others: R, S, EWMA (Exponentially Weighted Moving Average) [Roberts], CUSUM (CUmulative SUM) [Page], to give several examples, are just able to monitor one quality characteristic. These are univariate control charts. But, now, products quality is became a key of the companies' success, and they cannot approximate products quality with only one characteristic. So, multivariate control charts have been developed, and particularly the T^2 control chart [Hotelling].

These control charts are able to detect a fault (a disturbance) in a multivariate process. The fault can be a shift of the mean affecting one or more variables. Per example, it can be a step (sensor failure, operator change, and so on) or a trend (wearing tool, temperature variation, and so on). Major drawback of these multivariate control charts is that they give no indication about the root cause of a detected fault, and so they are not fully exploitable in the industry. In order to do the fault diagnosis of a detected fault, many approaches have been proposed. But, it will be of interest to have the possibility to detect and to diagnose a fault with an unique tool. An interesting approach for the diagnostic is the use of Bayesian Networks (BN) [Verron a, Verron b, Tiplica a]. In this article, we will study a possibility to detect a fault in a multivariate process with a BN. So, detection and diagnosis of a fault would be possible on a same tool: a bayesian network.

The article is structured as follows: in the second section we present utilization of multivariate control charts for the detection of faults on a multivariate process. The third section shows the approach that we propose to detect faults with a bayesian network. The bayesian network for detection is evaluated and compared to the T^2 and MEWMA multivariate control charts in the fourth section. In the last section, we conclude on the proposed approach and give outlooks.

2 – Multivariate control charts

2.1 – Hotelling T^2 chart

First works about the detection of fault in multivariate process began in 1947 with the Hotelling's works [Hotelling]. He was the first to propose a multivariate control chart based on a statistical distance. For a process with p variables, we can write the statistic T^2 as:

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

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 $p \times p$.

As we can see in the equation (1), the statistic T^2 is a scalar. So, we can plot the value of the T^2 for different instants, and with an appropriate control limit, we obtain the T^2 control chart. On this chart, each point represents the information of all the p variables. A fault is detected when a point is beyond the control limit.

As for the Shewhart chart, the set up of a T^2 chart is made in two phases. During the first phase, parameters of the process ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) are estimated. Concerning the computation of the parameters estimations and the control limit, readers can refer to the book of Montgomery [Montgomery]. In the case of individual measurements (sampling of unit size), Sullivan [Sullivan] does an interesting comparison between five estimators of $\boldsymbol{\Sigma}$. Once the parameters are estimated, the T^2 control chart can be drawn. It is very important to verify that the process is in control during this first phase. The second phase represents the real monitoring of the process on the assumption of a multivariate normal distribution with the estimated parameters.

2.2 – The *MEWMA control chart*

As the $\overline{X}$ control chart of Shewhart, the major drawback of the T^2 control chart is his moderate performances to detect shift of small magnitude. In order to cure this problem, others multivariate control charts have been proposed: MEWMA (Multivariate Exponentially Weighted Moving Average) [Lowry] and MCUSUM (Multivariate CUmulative SUM) [Pignatiello]. 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 of the process. So, it needs to do a transformation of the data process with the equation (2):

$$\mathbf{Y}_t = \boldsymbol{\lambda}\mathbf{X}_t + (\mathbf{I} - \boldsymbol{\lambda})\mathbf{Y}_{t-1} \quad (2)$$

where $\boldsymbol{\lambda}$ is a $p \times p$ diagonal weighting matrix, $\mathbf{I}$ is the $p \times p$ identity matrix, $\mathbf{X_t}$ is the observation vector (size $1 \times p$) at instant *t*, $\mathbf{Y}_0 = \boldsymbol{\mu}$ is the mean vector (size $1 \times p$) of the p variables.

Based on the same principle than a T^2 control chart, one can monitor the statistic of the equation (3):

$$T_t^2 = (\mathbf{Y}_t)^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} (\mathbf{Y}_t) \qquad (3)$$

where $\mathbf{Y}_t$ is the transformed observation vector at instant *t*, $\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}$ is the inverse of the variance-covariance matrix of $\mathbf{Y}_t$. It can be concluded that the process is out of control as soon as the T^2_t crosses a control limit h_M. Bodden [Bodden] proposed an algorithm to find the control limit h_M in order to respect a given number of false alarm and a given $\boldsymbol{\lambda}$.

In the last years, numerous methods have been proposed to find the root cause of a fault detected with a multivariate control chart [Mason, Chua]. It can be found a comparative study of different methods in [Tiplica b]. Without developing, we can say that these methods are based on different approaches: PCA (Principal Component Analysis), PLS (Projection to Latent Structure), decomposition of the Hotelling statistic, and so on. Of course, there is no easy response to this problem. The different methods are complex and ask a lot of time and computational resource in order to diagnose the cause of a fault. So, it seems to be an interesting approach to associate the fault detection and the fault diagnosis of a multivariate process in an only tool. Several works demonstrated that bayesian networks are able to diagnose correctly the fault of a multivariate process. So, it is of interest to study if bayesian network can do fault detection in an efficient way.

3 – Fault detection with bayesian network

3.1 – Bayesian networks

A bayesian network (BN), [Charniak] or [Naim] is a graph. In this graph, 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 {G, E, D} where:

{G} is a directed acyclic graph, G=(V,A), where V is the ensemble of nodes of G, and A is the ensemble of edges of G

{E} is a finite probabilistic space (Ω, Z, p)

{D} is an ensemble of random variables associated to the nodes of G and defined on E such as:

$$p(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} p(V_i | C(V_i)) \qquad (4)$$

where C(Vi) is the ensemble of causes (parents) of Vi in the graph G.

Inference in the network is based on the Bayes theorem:

$$p(X|Y)=\frac{p(Y|X)\times p(Y)}{p(X)} \qquad (5)$$

Bayesian network classifiers are particular bayesian networks [Friedman]. They always have a discrete node C coding the r different classes of the system. Other variables $X_1,\ldots, X_p$ represent the p descriptors (variables) of the system.

A famous bayesian classifier is the naïve bayesian classifier (or Bayes classifier) that we can see on the figure 1a. This bayesian classifier makes the strong assumption that descriptors of the system are class conditionally independent. Assuming normality of each descriptor, the naïve bayesian network is equivalent to the classification rule of the diagonal quadratic discriminant analysis. In order to alleviate to the assumption of independence, several approaches have been developed. We can cite the augmented naïve bayesian networks (figure 1b). These bayesian networks are based on a naïve bayesian networks but a tree is added between the descriptors. An other interesting approach is the Kononenko one [Kononenko], which represent some variables in one node (figure 1c). Assumption is made that this node follows a normal multivariate distribution (conditionally to the class).

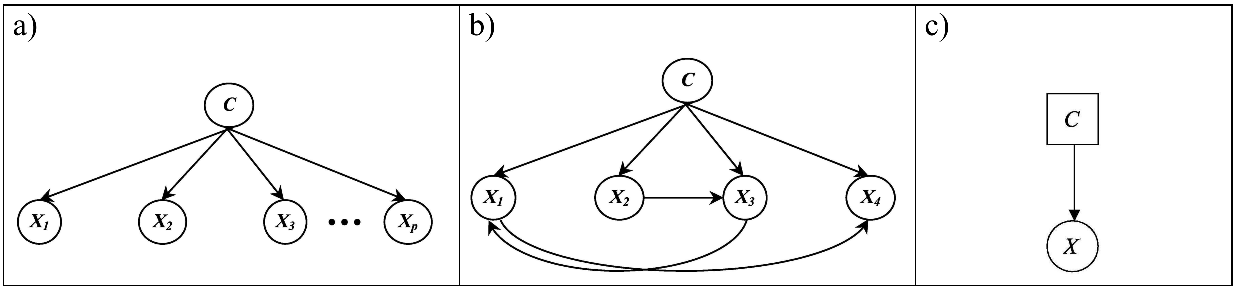

Figure 001: different bayesian network classifiers

3.2 – Structure used for the detection

The purpose of this article is to present a method allowing the fault detection in a multivariate process. The fault detection can be viewed as a classification task. Indeed, sensors data of the process represents values of the descriptors of the system. More, the system can be in two working classes: in control or out of control. The structure of the network will be composed of a class node (discrete variable) with two modalities (in or out of control). In order to represent the different descriptors of the system, we use an only multivariate variable. So, we obtain a structure similar to the one of the figure 1c.

In order to have a better adaptability of the network to different systems, we will assume that the multivariate data **Y** of the system are previously standardized. We remind that for multivariate data, standardization of the data is given by equation (6).

$$\mathbf{X}=\mathbf{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}) \qquad (6)$$

This equation (6) allows removing correlations between variables of the system. An example (on 2 dimensions) of this transformation is given on the figure 2.

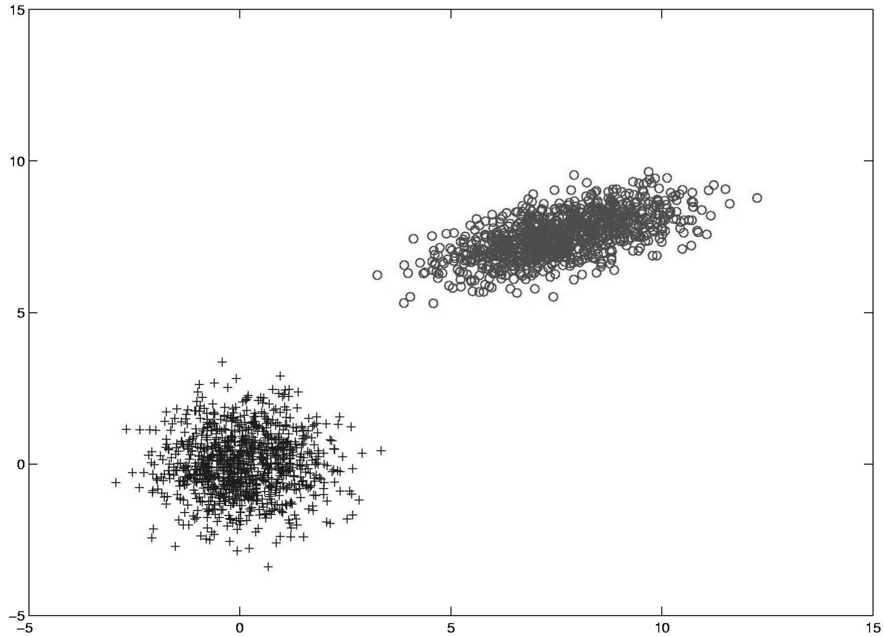

Figure 002: Example of the data transformation

Application of equation (6) allows to transform any normal multivariate data $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ in data $N(\mathbf{0},\boldsymbol{I})$ where $\mathbf{0}$ is the null vector of the same size than $\boldsymbol{\mu}$, and where $\boldsymbol{I}$ is the identity matrix of the same size than $\boldsymbol{\Sigma}$. Distribution of the "in control" state will be $N(\mathbf{0},\boldsymbol{I})$. In order to obtain a decision boundary between the two states of the process, we fix the distribution of the "out of control" state to $N(\mathbf{0},c\boldsymbol{I})$, where c is a coefficient strictly upper than 1. These two distributions will have same center ($\mathbf{0}$) and same shape (because $c\boldsymbol{I}$ is simply a scale extending of the shape of $\boldsymbol{I}$). The difference between the two states can be expressed as: scattering of the "out of control" state (HC) is more important than the scattering of the "in control" state (SC). So, we can represent the classification areas of the two states of the system for a bivariate example (see figure 3).

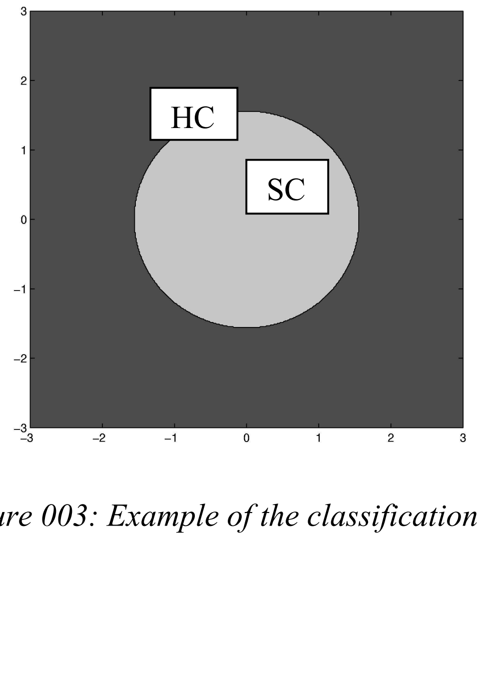

Figure 003: Example of the classification areas

On the figure 3, classifications areas represent area of the plan where probability of a modality (SC or HC) is the more important of the two. But, in order to correctly control the process, it is usual to accept (1-α)% of the population assumed to be “in control”, with α=0.01 or α=0.005. In the case of the figure 3, it is not the case. So, we must fix a threshold *na* allowing to not badly reject some situations “in control” (false alarms). This threshold is obtained by simulations, in order to satisfy a given number of false alarms. The structure of the network allows taking into account data of the system at instant *t*. But, it is very easy to add a multivariate node representing data of a required instant. So, the network is constructed in taking into account *h* instants: the present time *t*, and the *h*-1 last instants. So, the final structure of the network has a node class (C), linked to each multivariate node representing the system data to different instants (see figure 4).

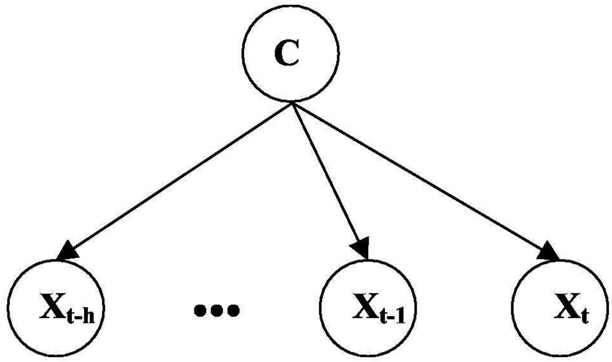

Figure 004: Final structure of the network

4 – Performances comparison of the different approaches

In order to evaluate the performances of our detection method with bayesian network, and to compare them to others detection methods (T^2 and MEWMA control charts), we have chosen simulation data. Performances of the different methods are evaluated with the ARL (Average Run Length). The ARL represents the average of the number of sample (of measurement) to take of the process between the beginning of a fault and his detection. The principle of the simulations is quite simple, for each case we have simulated 500 signals (length of 5000 samples) following a normal multivariate law of dimension *p*. The first *h* samples of the signal are “in control”, but from the sample *h*+1, a step of magnitude *k* is introduced in the signal. We have distinguished two types of step. Firstly, the step is only on one variable, in this case the magnitude of the step is *k* on the selected variable. Secondly, the step is on *p* variables. In this case, the step magnitude *k* is distributed on each variable as a step of magnitude $\frac{k}{\sqrt{p}}$. For each simulation, α is fixed to 0.005. For the MEWMA control charts, in order to correctly detect lower magnitude step all in keeping a reasonable ARL for large magnitude steps, we have chosen a λ equal to 0.4.

Table 001: ARL for p=2, h=5 and step on one variable

k	0	0.5	1	1.5	2	2.5	3	3.5	4
T^2	194.986	115.688	42.982	15.646	6.504	3.566	2.132	1.54	1.214
BN	200.856	89.644	24.582	8.642	4.212	2.904	2.1	1.752	1.414
MEWMA	198.672	53.63	13.294	5.828	3.29	2.516	1.916	1.668	1.45

Table 002: ARL for p=2, h=10 and step on one variable

k	0	0.5	1	1.5	2	2.5	3	3.5	4
T^2	202.656	124.746	42.924	16.762	6.568	3.728	2.294	1.496	1.21
BN	216.224	81.426	21.972	8.212	4.63	3.382	2.478	1.978	1.614
MEWMA	214.262	51.058	13.222	5.44	3.366	2.452	2.026	1.69	1.476

Table 003: ARL for p=10, h=10 and step on one variable

k	0	0.5	1	1.5	2	2.5	3	4	5
T^2	194.986	115.688	42.982	15.646	6.504	3.566	2.132	1.54	1.214
BN	200.856	89.644	24.582	8.642	4.212	2.904	2.1	1.752	1.414
MEWMA	198.672	53.63	13.294	5.828	3.29	2.516	1.916	1.668	1.45

Table 004: ARL for p=2, h=5 and step on p variables

k	0	0.5	1	1.5	2	2.5	3	4	5
T^2	192.496	113.58	40.25	16.52	6.56	3.736	2.08	1.48	1.22
BN	191.332	88.352	25.532	8.614	4.146	2.926	2.1	1.656	1.392
MEWMA	192.89	51.924	12.62	5.762	3.314	2.494	1.944	1.68	1.416

In looking at the 4 tables, we can see that the MEWMA control chart is the method giving the best results. But, the proposed method (BN), based on a bayesian network, obtains better results than a T^2 control chart. So, the proposed method allows detecting a step in a multivariate process.

Tables 1 and 2 show the influence of the parameter *h*. We can see that more *h* is low, more the larger magnitude steps will be rapidly detected. But, more *h* is high, more the lower magnitude steps are rapidly detected. So, to correctly use this method, *h* must be chosen in function of the step magnitude to be detected (in the same way that λ for a MEWMA control chart).

The results of tables 2 and 3 show the influence of the number *p* of the variable of the process. We can see that the method gives better performances with high step than with low step. So, performances of the bayesian network approach are not only function of *h*, but also function of the number of variables.

Finally, comparing table 1 and 4 show that our approach detects univariate (just on one variable) and multivariate (on multiple variable) steps. But, in order to improve performances of the bayesian network approach, it will be of interest to search informative variables of the system as proposed in [Verron b].

5 – Conclusions and outlooks

In this article, we have proposed a new approach allowing the fault detection in a multivariate process. This approach is based on a bayesian network. A class node represents the two states of the system ("in control" or "out of control"), and other nodes represent system data on multiple instants. Performances (in term of ARL) have been evaluated on simulation data and compared to MEWMA and T^2 control charts. We have seen that, on simulation data, performances of the bayesian network approach are lower than the MEWMA control chart, but performances of our approach outperforms the T^2 control chart.

The principal outlook to this work is the combination of the fault detection and the fault diagnosis in an only bayesian network.

Références :

[Bodden] Bodden, Rigdon. *A program for approximating the in-control ARL for the MEWMA chart*. Journal of Quality Technology, vol. 31, p120-123. 1999.

[Charniak] Charniak. *Bayesian networks without tears*. AI Magazine, vol.12, p50-63. 1991.

[Chua] Chua, Montgomery. *Investigation and Characterization of a Control Scheme for Multivariate Quality Control*. Quality and Reliability Engineering International, vol.8, p37-44. 1992.

[Hotelling] Hotelling. *Multivariate Quality Control*. Techniques of Statistical Analysis, p111-184. 1947.

[Kononenko] Kononenko. *Semi-naive Bayesian classifier*. EWSL-91: Proceedings of the European working session on learning on Machine learning, p206-219. Porto, 1991.

[Lowry] Lowry, Woodall, Champ, Rigdon. *A multivariate exponentially weighted moving average control chart*. Technometrics, vol.34, p46-53. 1992.

[Mason] Mason, Tracy, Young. *Decomposition of T2 for multivariate control chart interpretation.* Journal of Quality Technology vol.27, p99-108. 1995.

[Montgomery] Montgomery. *Introduction to Statistical Quality Control Third Edition*. John Wiley and Sons. 1997.

[Naim] Naim, Wuillemin, Leray, Pourret, Becker. *Réseaux Bayésiens - 2ème édition*. Eyrolles. 2004.

[Page] Page. *Continuous Inspection Schemes*. Biometrika, vol.41, p100-115. 1954.

[Pignatiello] Pignatiello, Runger. *Comparisons of Multivariate CUSUM Charts*. Journal of Quality Technology, vol.22, p173-186. 1990.

[Roberts] Roberts. *Control chart tests based on geometric moving averages*. Technometrics, vol.1, p239-250. 1959.

[Shewhart] Shewhart. *Economic control of quality of manufactured product*. New York : D. Van Nostrand Co. 1931.

[Sullivan] Sullivan, Woodall. *A Comparison of Multivariate Control Charts for Individual Observations*. Journal of Quality Technology, vol.28, p398-408. 1996.

[Tiplica a] Tiplica, Verron, Kobi, Nastac. *FDI in multivariate process with naïve bayesian network in the space of discriminant factors*. Accepted In International Conference on Computational Intelligence for Modelling, Control and Automation, Sydney, Australia, 2006.

[Tiplica b] Tiplica, Kobi, Barreau. *Synthèse et Comparaison des Méthodes pour la Maîtrise Statistique des Processus Multivariés*. Actes du congès QUALITA, p134-142. Annecy, France, 2001.

[Verron a] Verron, Tiplica, Kobi. *Détection des fautes dans un procédé multivarié et aide au diagnostic par réseaux bayésiens naïfs.* In Conférence Internationale en Métrologie, Casablanca, Maroc, 2006.

[Verron b] Verron, Tiplica, Kobi. *Fault diagnosis with bayesian networks: Application to the tennessee eastman process*. Accepted In IEEE International Conference on Industrial Technology, Mumbai, India, 2006.