Bayesian networks and mutual information for fault diagnosis of industrial systems

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi

BAYESIAN NETWORKS AND MUTUAL INFORMATION FOR FAULT DIAGNOSIS OF INDUSTRIAL SYSTEMS

Sylvain Verron [*,1] **Teodor Tiplica** [*]
Abdessamad Kobi [*]

[*] *LASQUO/ISTIA, University of Angers, 49000 Angers, France*

Abstract: The purpose of this article is to present and evaluate the performance of a new procedure for industrial process diagnosis. This method is based on the use of a bayesian network as a classifier. But, as the classification performances are not very efficient in the space described by all variables of the process, an identification of important variables is made. This feature selection is made by computing the mutual information between each process variable and the class variable. The performances of this method are evaluated on the data of a benchmark problem: the Tennessee Eastman Process. Three kinds of faults are taken into account on this complex process. The objective is to obtain the minimal recognition error rate for these 3 faults. Results are given and compared with results of other authors on the same data.

Keywords: Fault diagnosis, bayesian network classfiers

1. INTRODUCTION

Nowadays, industrial processes are more and more complex. So, they include a lot of sensors. As a consequence, an important amount of data can be obtained from a process. A process dealing with a lot of variables can be named multivariate process. But, the monitoring of a multivariate process cannot be reduced to the monitoring of each process variable because the correlations between the variables have to be taken into account. Process monitoring is an essential task. The final goal of the process monitoring is to reduce variability, and so, to improve the quality of the product (Montgomery, 1997). The process monitoring comprises four procedures: fault detection (decide if the process is under normal condition or out-of-control); fault identification (identify the variables implicated in an observed out-of-control status); fault diagnosis (find the root cause of the disturbance); process recovery (return the process to a normal status).

Three major kinds of approaches exists for process monitoring (Chiang *et al.*, 2001): data-driven, analytical and knowledge-based. Theoretically, the best method is the analytical one because this method constructs mathematic models of the process. But, for large systems (lots of inputs, outputs and states), obtaining detailed models is almost impossible. In the knowledge-based category are placed methods that are based on qualitative models and methods that are black boxes (like neural networks). Finally, data-driven methods are techniques based on rigorous statistical development of process data. Our interest is to monitor large systems, and so, we are concerned with data-driven methods.

In literature, we can find a lot of different data-driven techniques for process control. For the fault detection of industrial processes many methods have been submitted: univariate statistical process control (Shewhart charts) (Shewhart, 1931; Roberts, 1959), multivariate statistical process control (T^2 and Q charts) (Hotelling, 1947; Westerhuis *et al.*, 2000), and some PCA (Principal Component Analysis) based techniques (Jackson, 1985) like Multiway PCA or Moving PCA (Bakshi, 1998). An interesting article is (Kano *et al.*, 2002) in which authors make comparisons between these different techniques. For the fault identification procedure, one of the better statistical techniques is the MYT decomposition of the T^2 statistic (Mason *et al.*, 1995; Mason *et al.*, 1997). Finally, for the fault diagnosis techniques we can cite the book of Chiang, Russell and Braatz (Chiang *et al.*, 2001) which presents a lot of them (PCA based techniques, Fisher Discriminant Analysis, PLS based techniques, etc).

In this article, we present a new procedure for the diagnosis of the faults of an industrial system based on the utilization of mutual information and bayesian network. The mutual information is used for the selection of the key variables of the system, and bayesian networks are used for classification in the space of the key variables. The article is structured in the following manner. In the second section, we introduce bayesian network (BN) classifiers and some notions concerning the mutual information. In the third section, we present the new diagnosis procedure that we are proposing, using bayesian network and mutual information, to diagnosis the fault of an industrial system. The fourth section presents an application of this procedure for the diagnosis of three types of fault on the benchmark problem Tennessee Eastman Process. Finally, in the last section, we conclude on the advantages and drawbacks of this procedure, and we present outlooks of fault diagnosis with bayesian network classifiers.

2. TOOLS

2.1 Bayesian network classifiers

A Bayesian Network (BN) (Jensen, 1996) is an acyclic graph in which each variable is a node that can be continuous or discrete. Edges of the graph represent dependence between linked nodes.

Bayesian network classifiers are particular bayesian networks (Friedman *et al.*, 1997). 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 Naïve Bayesian Networks (NBN) is particular type of BN classifiers (Friedman *et al.*, 1997).

It is also known as the Bayes classifier. A NBN is composed of a "class node" representing a variable which modalities are the different classes of our system. This node is thus linked with all other variables of the system (descriptors) as indicated on the figure 1.
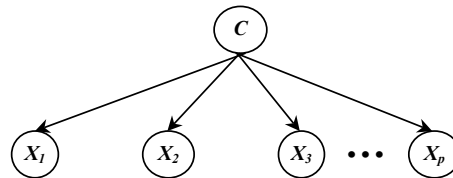

Fig. 1. Naïve bayesian network

The NBN is called naïve because it makes the naïve (but strong) assumption that all descriptors (variables of the system) are statistically independent (no correlation between each other). But, this classifier is not well fitted for industrial systems. Indeed, in industrial 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. 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. So, some normally distributed variables can be modeled with a node representing a multivariate normal distribution. In that 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 2.
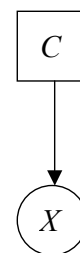
Fig. 2. Condensed Semi Naïve Bayesian Network

Although this classifier is well fitted for the classification task of faults in industrial, it still remains a problem. If we have non-informative descriptors, the performances (in term of correct classification rate) are poor. So, if we want to diagnosis a system which has many variables, even though only few are really important for classification, other (less important) must not be taken into account. We have to do a selection of the important variables for classification. We will use mutual information for representing the importance of a variable in the classification (discrimination) task.

2.2 Mutual Information

In the information theory (Shannon, 1948), the Mutual Information (I) of two random variables is a quantity measuring the mutual dependence of the two variables. In our case, the objective is to find important descriptors as well as to quantify this importance. For that, we must compute, for $i = 1, 2, \ldots, p$, $I(C; X_i)$ the mutual information between the class variable C and each descriptor X_i.

In (Perez *et al.*, 2006), authors demonstrate that the mutual information between a gaussian (normaly distributed) variable and a multinomial (discrete) variable can be computed as indicated by equation 1. For this equation, it is assumed that: C is a multinomial random variable with r possible values and a probability distribution given by $P(C = c) = P(c)$; X is a random variable with a normal density function of parameters μ et σ^2; X conditioned to $C = c$ follows a normal density function with parameters μ_c et $\sigma_c{}^2$.

$$I(X; C) = \frac{1}{2}\left[\log(\sigma^2) - \sum_{c=1}^{r} P(c)\log(\sigma_c^2)\right] \quad (1)$$

This equation can be reduced on the form of the equation 2.

$$I(X; C) = \log(\sigma) - \sum_{c=1}^{r} P(c)\log(\sigma_c) \quad (2)$$

So, the mutual information (I) can be computed for all variables (descriptors) of the industrial system. The most important variables for the classification task will be those having an important I value.

We have seen a tool able to classify (CSNBN) and a tool able to quantify the discriminant information of each descriptor (mutual information I). Now, we will see how to use these two tools for diagnosis a fault of an industrial system.

3. NEW PROCEDURE FOR FAULT DIAGNOSIS

The main idea of the procedure we are proposed for fault diagnosis in industrial processes with bayesian network classifier is described on figure 3.

This procedure requires a faults database with examples of each fault that we want to diagnosis. The mutual information between class node C (variable which codes the type of fault) and each descriptor X_i (variable of the process) can be
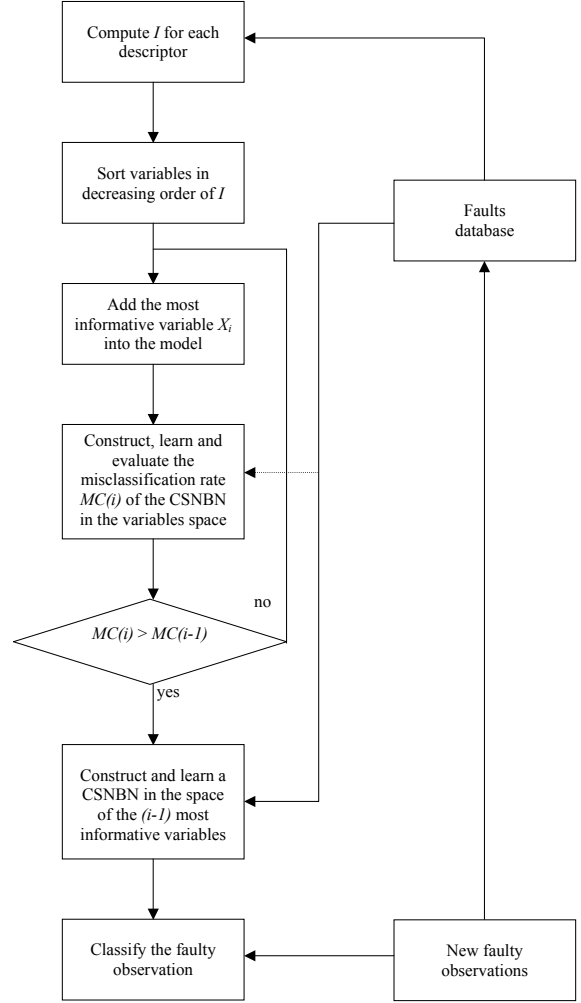

Fig. 3. Fault diagnosis procedure

computed. So, we can sort the variables in decreasing order of mutual information. After that, we will iteratively add variables to the CSNBN and evaluate the misclassification rate. First we will observe that the misclassification rate decreases by adding variables into the model. But once a certain number of variables is overpassed, the misclassification rate starts to increase again. If the misclassification rate increases at step i, the best model will be the one with the $(i - 1)$ first informative variables.

Of course, the evaluation of classifiers is done only on the training dataset. For this evaluation procedure, we apply a well known technique: the m-fold cross validation In the m-fold cross validation, the training dataset is divided into m subsets. So, for each m, one of the m subsets is used as the testing set, and the m-1 other subsets are put together to form the training set. Then the average error across all m trials is computed (Duda *et al.*, 2001).

Now, we will see an application of this approach on a benchmark problem: the Tennessee Eastman Process (figure 4).
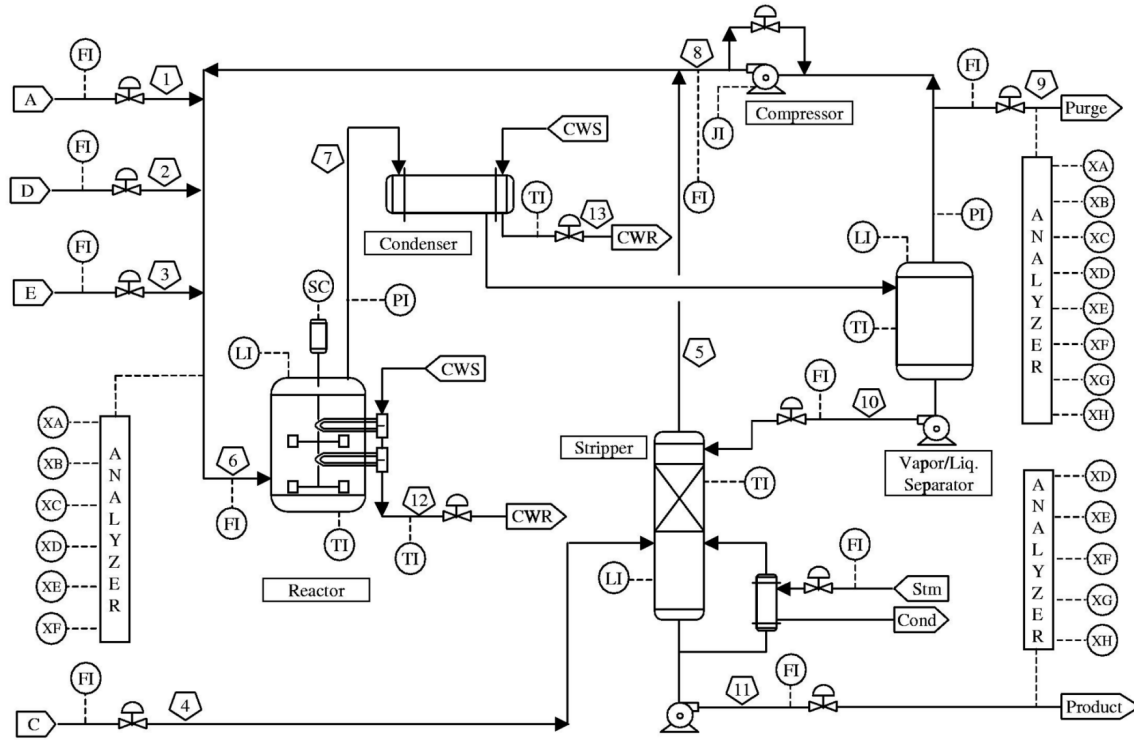
Fig. 4. Process flowsheet of the TEP

4. APPLICATION TO THE TEP

4.1 Presentation of the TEP

The Tennessee Eastman Process (TEP) is a chemical process. It is not a real process but a simulation of a process that was created by the Eastman Chemical Company to provide a realistic industrial process in order to evaluate process control and monitoring methods. Article of Downs and Vogel (Downs and Vogel, 1993) entirely describes this process. Authors also give the Fortran code of the simulation of the process. Ricker (Ricker, 1996) has implemented the simulation on Matlab. The TEP is composed of five major operation units: a reactor, a condenser, a compressor, a stripper and a separator. Four gaseous reactant A, C, D, E and an inert B are fed to the reactor where the liquid products F, G and H are formed. This process has 12 input variables and 41 output variables. The TEP has 20 types of identified faults. So, this process is ideal to test monitoring methods. But, it is also a benchmark problem for control techniques because it is open-loop unstable. Many articles (authors) present the TEP and test their approaches on it. For example, in fault detection, we can cite (Kano *et al.*, 2002; Lee *et al.*, 2004; Kruger *et al.*, 2004). Some fault diagnosis techniques have also been tested on the TEP (Chiang *et al.*, 2001; Chiang *et al.*, 2004; Kulkarni *et al.*, 2005) with the plant-wide control structure recommended in Lyman and Georgakis (Lyman and Georgakis, 1995). In (Chiang *et al.*, 2004; Kulkarni *et al.*, 2005), au-

Class	Fault type	Train data	Test data
1	Fault 4: step change in the reactor cooling water inlet temperature	480	800
2	Fault 9: random variation in D feed temperature	480	800
3	Fault 11: random variation in the reactor cooling water inlet temperature	480	800

Table 1. Description of fault datasets

thors focus on only 3 types of fault and give the datasets they used. For this reason, we will take the same data that in these articles and compare our approach to those of the others.

As we said, we have taken into account 3 types of faults: fault 4, 9 and 11 (see table 1). These three types of faults are good representations of overlapping data and so, are not easy to classify.

As indicated on the table 1, each type of fault have 2 datasets: a training sample and a testing sample, containing respectively 480 and 800 observations.

4.2 Approach on the TEP

Now, we will apply our approach on the TEP application. We first compute I for all variables and sort variables in decreasing order of I. After that, we iteratively add variables to the model and

compute misclassification rate of a 10-fold cross validation on the training dataset. Result of these misclassification rates are given in table 2.

Variable(s) in the model	Misclassification rate
51	12.5
51, 9	6.46
51, 9, 50	6.04
51, 9, 50, 19	5.90
51, 9, 50, 19, 18	**5.62**
51, 9, 50, 19, 18, 20	5.70

Table 2. Misclassifications of steps

On this table 2, we can see that best result (on training dataset) is obtained for the selection of the 5 first informative variables (variables 51, 9, 50, 19 and 18). So, we construct a CSNBN in the space of these determined key variables and learn parameters of the CSNBN by MLE. Finally, we can classify the 2400 observations of the testing dataset. Results of this classification can be seen on the confusion matrix of the table 3.

Class	Fault 4	Fault 9	Fault 11	Total
Fault 4	788	0	31	819
Fault 9	0	778	74	852
Fault 11	12	22	695	729
Total	800	800	800	2400

Table 3. Confusion matrix

Based on this table, we can compute the misclassification rate. Its value is 5.79%. We can now analyze our results and compare them to those of other authors.

4.3 Analysis and comparison of results

We will compare results of our approach to those of other methods on the same data (see table 4). We notify that results for methods FDA (Fisher Discriminant Analysis), SVM (Support Vector Machines) and PSVM (Proximal Support Vector Machines) are extracted from (Chiang et al., 2004), and results for ISVM (Independent Support Vector Machines) from (Kulkarni et al., 2005). In these articles, classification was made in the space of the variables 9 and 51. In the table 4, we have added the misclassification rate of the CSNBN in the space of all the variables (53 variables) $CSNBN_{all}$. We have also computed the misclassification rate that would be obtained by the use of NBN, firstly in the space of all the variables (NBN_{all}), secondly in the space described by variables 9 and 51 ($NBN_{9,51}$) and finally in the same space that the result of our approach ($NBN_{9,18,19,50,51}$).

We can see that the classification methods in the selected variables space give better results than in the space of all the variables. Second interesting remark is the fact that on the same space (variables 9 and 51), the best classifier is the CSNBN.

Method	Misclassification rate
NBN_{all}	20.83%
$CSNBN_{all}$	18.83%
$FDA_{9,51}$	16%
$PSVM_{9,51}$	6.0%
$SVM_{9,51}$	6.5%
$ISVM_{9,51}$	6.0%
$NBN_{9,51}$	7.79%
$CSNBN_{9,51}$	5.87%
$NBN_{9,18,19,50,51}$	13.12%
$CSNBN_{9,18,19,50,51}$	**5.79%**

Table 4. Results of the application

Indeed, $CSNBN_{9,51}$ outperforms all other classifiers (FDA, NBN and SVM based classifiers). The fact that CSNBN outperforms FDA is not surprising because FDA is a linear technique while CSNBN is a quadratic one. More, it is also logic than CSNBN outperforms NBN because NBN do not take into account correlation between variable 9 and variable 51 (correlation of 0.42). Although CSNBN requires quite simple computation, its results are quite similar to SVM based techniques (SVM, PSVM, ISVM), which are techniques requiring more computational potential. Finally, we can see that our method gives the best misclassification rate for this example of the TEP. But, it will be interesting to have misclassification rates of others techniques in the same space of variables (9, 18, 19, 50, 51).

Our approach gives good results on this example because fault classes are normally distributed (ellipsoidal shapes). Results of the CSNBN would not be so efficient if the data was not normal. But, an interesting way for taking into account not normally distributed variable is gaussian mixture model which, on the based of few gaussian distributions, can represent any type of distribution. It will be a good answer to the problem of normally distributed data assumption. More, it is easy to represent gaussian mixture model with a bayesian network.

Another way of interest is the diagnosis of a not identified fault type (no historical data on this fault type). Indeed, major drawback of our procedure is the selection of variables able to discriminate identified faults. But, if a new type of fault appeared, not discriminable on the selected variables, procedures randomly attributes to this fault an identified type.

5. CONCLUSION AND OUTLOOKS

The main interest of this article is the presentation of a new procedures for fault diagnosis of industrial processes. This procedures uses a faults database to compute mutual information between each descriptor and the fault type variable. At this time, informative variables are iteratively added to the model and misclassification rates are estimated with a 10-fold cross validation. Iteration stops as soon as the estimated misclassification rate increases. Performances of this approach was tested on a concrete example: the Tennessee Eastman Process. Results of this method are good and outperform some previous results of other published methods.

Outlooks of these procedures will be application of similar techniques, based on bayesian networks, in order to improve the fault diagnosis (cases of a non identified fault and non normally distributed data), but also in order to study in which way the fault detection step can be made with bayesian networks. Final goal is to have on the same procedure the fault detection and the fault diagnosis.

REFERENCES

Bakshi, Bhavik R. (1998). Multiscale pca with application to multivariate statistical process monitoring. *AIChE Journal* **44**(7), 1596–1610.

Chiang, Leo H., Evan L. Russell and Richard D. Braatz (2001). *Fault detection and diagnosis in industrial systems*. New York: Springer-Verlag.

Chiang, L.H., M.E. Kotanchek and A.K. Kordon (2004). Fault diagnosis based on fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering* **28**(8), 1389–1401.

Downs, J.J. and E.F. Vogel (1993). Plant-wide industrial process control problem. *Computers and Chemical Engineering* **17**(3), 245–255.

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

Friedman, N., D. Geiger and M. Goldszmidt (1997). Bayesian network classifiers. *Machine Learning* **29**(2-3), 131–163.

Hotelling, Harold (1947). Multivariate quality control. *Techniques of Statistical Analysis* pp. 111–184.

Jackson, Edward J. (1985). Multivariate quality control. *Communication Statistics - Theory and Methods* **14**, 2657 – 2688.

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

Kano, M., K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss and B.R. Bakshi (2002).

Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Computers and Chemical Engineering* **26**(2), 161–174.

Kononenko, Igor (1991). Semi-naive bayesian classifier. In: *EWSL-91: Proceedings of the European working session on learning on Machine learning*. pp. 206–219.

Kruger, U., Y. Zhou and G.W. Irwin (2004). Improved principal component monitoring of large-scale processes. *Journal of Process Control* **14**(8), 879–888.

Kulkarni, A., V.K. Jayaraman and B.D. Kulkarni (2005). Knowledge incorporated support vector machines to detect faults in tennessee eastman process. *Computers and Chemical Engineering* **29**(10), 2128–2133.

Lee, J.-M., C.K. Yoo and I.-B. Lee (2004). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science* **59**(14), 2995–3006.

Lyman, P.R. and C. Georgakis (1995). Plant-wide control of the tennessee eastman problem. *Computers and Chemical Engineering* **19**(3), 321–331.

Mason, R.L., N.D. Tracy and J.C. Young (1997). A practical approach for interpreting multivariate t2 control chart signals. *Journal of Quality Technology* **29**(4), 396–406.

Mason, Robert L., Nola D. Tracy and John C. Young (1995). Decomposition of t2 for multivariate control chart interpretation. *Journal of Quality Technology* **27**(2), 99–108.

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

Perez, Aritz, Pedro Larranaga and Inaki Inza (2006). Supervised classification with conditional gaussian networks: Increasing the structure complexity from naive bayes. *International Journal of Approximate Reasoning* **In Press, Corrected Proof**, –.

Ricker, N.L. (1996). Decentralized control of the tennessee eastman challenge process. *Journal of Process Control* **6**(4), 205–221.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1**(3), 239–250.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.* **27**, 379–423, 623–656.

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

Westerhuis, J.A., S.P. Gurden and A.K. Smilde (2000). Standardized q-statistic for improved sensitivity in the monitoring of residuals in mspc. *Journal of Chemometrics* **14**(4), 335–349.