Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Latent and small fault detection and diagnosis for dynamic processes

Zhiqiang Ge, Zhihuan Song

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, China
e-mail: zqge@iipc.zju.edu.cn, zhsong@iipc.zju.edu.cn

Abstract: Compared to large process faults, the latent and small ones are difficult to be detected. However, the accumulation of these faults may even more harmful to the process. A novel fault detection and diagnosis method is proposed which is based on similarity factor and a variable moving window. The new method is based on the idea that a change of process can be reflected in the distribution of the data, which can be detected more easily by the proposed similarity factor. Meanwhile, it has no Gaussian distribution limitation of the process data, since the mixed similarity factor is introduced. The independent component analysis (ICA) factor and the principal component analysis (PCA) factor are used for similarity comparison for Gaussian and non-Gaussian information, respectively. Besides, in order to determine the dynamic step accurately and cut the computation cost, the conventional dynamic method is modified by using autocorrelation analysis. A case study of Tennessee Eastman (TE) benchmark process shows the efficiency of the new proposed method.

1. INTRODUCTION

In order to monitor multivariable processes, multivariate statistical process control (MSPC) has been developed. In the last decade, various extensions of MSPC have been proposed and widely used in industrial processes. [Kano, et al., 2002b] However, conventional MSPC methods do not always function very well, because they can not detect the change of correlation among process variables as long as both of the statistics are inside their corresponding control limits. These process changes are latent and small, in some cases they can be easily compensated by the control systems. Therefore, they are difficult to be detected by conventional methods. However, catastrophic consequences can result from the accumulation of these latent and small faults. Besides, for dynamic processes, the dynamic step for the conventional dynamic principal component analysis (DPCA) [Ku, 1995] is difficult to be determined, and the large data matrix involves great computation burden.

In recently years, a new monitoring index known as DISSIM was proposed by Kano, et al. [2002a], which was based on the dissimilarity analysis of process data. compared to the conventional MSPC method, this method is based on the idea that a change of operating condition can be reflected in the distribution of the process data, which can be detected by the proposed DISSIM method. Several successful theory researches and applications have demonstrated that the method can quickly and effectively detect the change of correlations among process variables. [Kano, et al., 2002b; Zhao, et al., 2007] Another type of novel methods is the similarity factor based method which is used to identify the similarity between different operation modes. [Krzanowski, 1979; Singhai, and Seborg, 2006] However, the two types of methods are based on the assumption that the data formed Gaussian distribution. Ge, and Song [2007] proposed a new mixed similarity factor for fault mode identification. However, the fault detection method based on the proposed similarity factor has not been developed. Because the similarity factor is based on the distribution of the data, changes (latent or small) can be reflected in their corresponding data distributions. If the data distribution changes, the related process fault can be detected more easily.

In order to detect the latent and small faults in dynamic processes, the conventional dynamic method is modified. For online monitoring, the data matrix representing the current operating condition is updated by a variable moving window. After the latent or small fault has been detected, a new fault diagnosis method is proposed to identify the root cause. The rest of the paper is organized as follows. First, the conventional dynamic method is modified. Then the mixed similarity factor is introduced and the new fault detection and diagnosis method is proposed. In section 4, a case study of TE process is demonstrated. Conclusions are presented in the last section.

2. A MODIFIED APPROACH FOR DYNAMIC PROCESSES

In most dynamic cases, static fault detection methods do not function well for autocorrelated data. One useful approach is using past measurements as monitored variables for capturing the correlation among process variables, since the dynamic can be taken into accout. [Ku, 1995] The data matrix of process variables for dynamic monitoring is

$$X=[X(t-l)\ \ X(t-l+1)\cdots X(t-1)\ \ X(t)] \quad (1)$$

Where $X(t)$ is the current process data matrix, l is dynamic step, which is difficult to determine. It will involve large computation when the number of process variables is huge. Besides, it will also trigger excessive false alarms when unnecessary dynamic steps are taken for some process variables. In the present paper, the traditional dynamic monitoring method is modified. Thus autocorrelation analysis is introduced to determine the dynamic steps for each of the process variable. The autocorrelation coefficient is calculated as follow:

$$\rho(\tau) = \frac{\operatorname{cov}(X(t), X(t-\tau))}{\sqrt{D(X(t))} \cdot \sqrt{D(X(t-\tau))}} \tag{2}$$

Where cov is covariance of the two vectors, and D is variance of the corresponding vector. If $\rho(\tau)$ has a big value (but not bigger than 1), the correlation between the two vectors is strong. Otherwise, the correlation between the two vectors is weak if the value of $\rho(\tau)$ is small. A cut-off value could be set to determine whether the correlation between the considered vectors is significant. Therefore, variables with strong autocorrelation will take big dynamic steps, while small or no dynamic steps will be taken for variables with little and no autocorrelation. The data matrix of process variables for modified dynamic monitoring becomes:

$$[X_1(t-l_1)\cdots X_1(t) \;\; X_2(t-l_2)\cdots X_2(t) \;\; \cdots \;\; \cdots \;\; X_m(t-l_m)\cdots X_m(t)] \tag{3}$$

Where the sub-matrix $[X_i(t-l_i)\cdots X_i(t)]$ corresponds to the i-th process variable, and l_i is the dynamic step of the i-th variable.

3. FAULT DETECTION AND DIAGNOSIS METHOD BASED ON SIMILARITY FACTOR AND MOVING WINDOW

3.1 Similarity factor

In our previous work, a two-step ICA-PCA information extraction strategy was proposed, which is not under the assumption of Gaussian distribution of the process data. Instead, the non-Gaussian data information can also be extracted. Based upon this information extraction strategy, the mixed similarity factor was proposed. Krzanowski [1979] developed a method for measuring the similarity of two datasets using a PCA similarity factor S_{PCA}:

$$S_{PCA} = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}\cos^2\theta_{ij} \tag{4}$$

Where θ_{ij} is the angle between the *i*-th principal component of the first dataset and the j-th principal component of the second dataset, k is the number of retained principal components. The ICA similarity factor can be calculated from the main angles between the two ICA subspaces:

$$S_{ICA} = \frac{1}{r}\sum_{i=1}^{r}\cos^2\theta_i \tag{5}$$

Where θ_i is the *i*-th main angle between the two ICA subspaces, see Ge, and Song [2007]. Then the mixed similarity factor is defined as:

$$S_{mix} = \lambda \cdot S_{PCA} + (1-\lambda)\cdot S_{ICA} \tag{6}$$

Where $0 \le \lambda \le 1$. The choices of these factors depend on the feature of the process, and prior knowledge is also very useful. In the mixed similarity factor, both of the two kind of information (Gaussian and non-Gaussian) are considered. It has no distribution limitation and thus will be more reliable than the conventional one. However, there may exist such a case that two datasets have similar distribution directions but have different locations. In this case, the distance similarity factor should be incorporated, which was indicated in reference [Ge, and Song, 2007].

Since the similarity factors (S_{PCA}, S_{ICA} and S_{mix}) are all based on the distribution of the process data, the process change can be reflected in their model structure (PCA and ICA subspaces). Therefore, this method can detect changes in the directions of the two subspaces that span the model plane. Latent and small faults can also change the distribution of the process data, thus they can be detected by the proposed method. The case study of the present paper demonstrated that the proposed method shows good performance and sensitivity for latent and small faults monitoring.

3.2 Moving window approach

Generally, a constant length moving window is employed to track the process. However, it is always very important to select proper size for such a moving window, determination of the moving window size is still an open problem. In the present paper, a variable moving window is used [Zhao, et al., 2007]. The moving window length grows gradually with the process going on, thus more valuable data will be consisted in the moving window. However, during the early period of process, because there is not enough sample information, the moving window may not sufficiently reflect the current process operating condition. Besides, the changes of the independent subspace and principal subspace will be too sensitive, which may cause inevitable false alarms. To solve these problems, an initial window is introduced, which contains normal operating sampling data. Therefore, the moving window grows gradually based on the initial window. Then with the development of process, new sample information is added step by step, thus the moving window becomes longer and longer. If the window size becomes bigger than the size of the reference window, the window is cut off to represent the most recent information as the same size of the reference window. With this method, the choice of moving window size will not impose great influence on the monitoring effect any more. The choice of the initial window size also shows no great influence on the monitoring effect, therefore can be chosen by experience.

3.3 Fault detection method

For continuous process modelling, using similarity factor and moving window, the reference data under normal operating condition should be defined previously. Then the similarity between the moving window and the reference data set can be calculated. For monitoring purpose, it is better to convert similarity to dissimilarity. The conversion is defined as follow:

$$DS_{mix} = 1 - S_{mix} \tag{7}$$

After the comparison between normal process data and the reference data set, a proper control limit should be determined for process monitoring. In DISSIM method, [Kano, et al., 2002a] the control limits 99% and 95% are determined simply so that the ratio of samples outside the limits to the entire samples is 1% and 5% respectively. Recently in EDISSIM method, Zhao, et al. [2007] assumed that the distribution of dissimilarity factor follows Γ - distribution. However, this is not always true due to the different kind of process and operating conditions. In the present paper, the control limit is defined by a non-parametric empirical density estimates using kernel extraction.[Chen, et al., 2004] One major advantage of the control limit obtained using kernel density estimation is that it follows the data more closely, and is less likely to incorporate regions of unknown operation than the one obtained from the conventional methods.

The illustration and detail of the proposed fault detection method is shown as follows:

Modelling phase:

Step 1: Acquire time-serial data when then process is operated under normal condition. Determined the dynamic step for each variable, and normalize each variable with its mean and standard deviation values;

Step 2: Select size of the initial window, choose a reference data set;

Step 3: Generate data set with the variable moving window step by step, they are scaled with the mean and the variance obtained at step 1;

Step 4: Build ICA-PCA model for these moving windows and the reference data set, then the similarity factor S_{mix} are calculated at each step, and converted to the dissimilarity factor DS_{mix};

Step 5: Determine the control limit for the dissimilarity factor DS_{mix} by kernel density estimation.

Online fault detection phase:

Step 1: Acquire time-serial data by the moving window, and normalization with the same mean and standard deviation obtained from the step 1 of modelling phase. Construct new data vector for dynamic monitoring based on dynamic step of each variable.

Step 2: Calculate the PCA, ICA, and mixed similarity factors between the new dataset and the conference dataset.

Step 5: If the value of DS_{mix} rejects its corresponding control limit, then a fault alarm will be trigged. If the value is inside the control limit, keeping monitoring.

3.4 Fault diagnosis method

Typically, the contribution plots are used to identify the root cause of the abnormal operation. However, they can only identify some simple faults, and based on T^2 and SPE statistics. Since these statistics are not used for fault detection in the present work, the contribution plots are hardly to be carried out. Instead, the dissimilarity factor DS_{mix} is used for fault detection. Therefore, a new fault diagnosis tool is developed based on these factors, which is called sub-dissimilarity factor (SubDS) in the present paper. The principle of the new proposed fault diagnosis tool is demonstrated as follow: when a fault is detected, assume some process variable contributes significantly to this fault. Thus the dissimilarity factor goes beyond the bound of normal operation due to this responsible variable. If the responsible variable is removed from both of the moving window and the reference data set, then the new calculated value of the sub-dissimilarity factor should be below the corresponding control limit. Otherwise, if any other process variable is removed, the new calculated value of the sub-dissimilarity factor will continue to stay above the control limit. However, there may be several variables responsible for the fault, which always true in practice. In this case, the sub-dissimilarity factor could be ranged by its value. The variable corresponding to the smallest value of SubDS should be the most likely root cause of the fault. Thus,

$$\begin{aligned} &SubDS(v) = \min\{SubDS(j)\} \\ &j = 1, 2, \cdots m \end{aligned} \tag{8}$$

Where v is considered the responsible variable of the fault, m is the number of the variables for monitoring.

The details of the proposed fault diagnosis procedures are given as follow:

Step 1: When some fault is detected, thus the dissimilarity factor DS_{mix} goes beyond the control limit. Choose the current moving window and the reference data set for fault diagnosis;

Step 2: Remove one variable from the moving window and the reference data set every time, calculated the new dissimilarity factors.

Step 3: Range the values of SubDS, the variable corresponding to the smallest value of SubDS should be judged to be responsible for the detected fault.

The process of fault diagnosis is shown in Fig. 1.

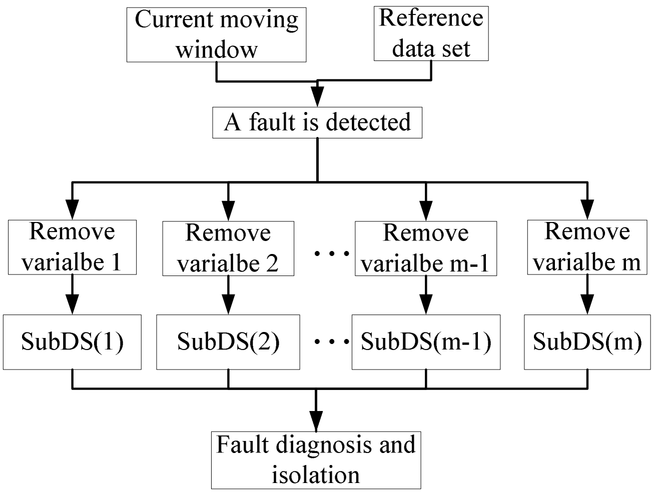


Fig. 1. Flowsheet of fault diagnosis method

4. CASE STUDY OF TE PROCESS

The TE benchmark process has been widely used to test the performance of various fault detection method. [Chiang, et al., 2000] This process consists of five major unit operations: a reactor, a condenser, a compressor, a separator, and a stripper. The control structure is shown schematically in Fig. 2, which is the second structure listed in Lyman and Georgakis [1995]. The TE process has 41 measured variables and 12 manipulated variables, a set of 21 programmed faults are introduced to the process.

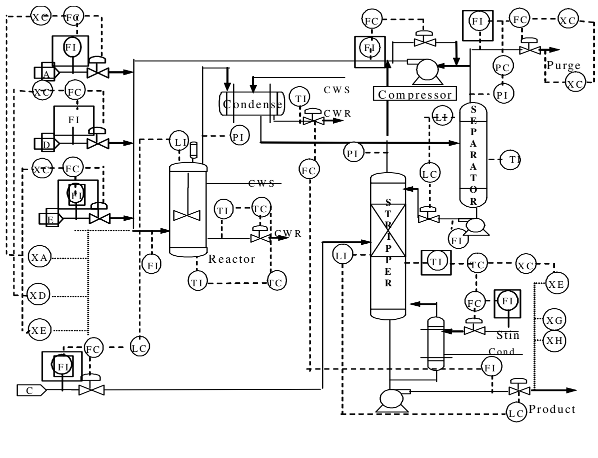


Fig. 2. Control system of the Tennessee Eastman process

In the present paper, the variables selected for monitoring are listed in Table 1, which are the same as Kano et al. [2002a] There are 16 variables selected for monitoring. The simulation data which we have collected were separated into two parts: the training datasets and the testing datasets, they consisted of 960 observations for each operating mode (1 normal and 21 fault), respectively, and their sampling interval was 3 min. Among these faults, some faults are easy to be detected, since they greatly affect the process and change the relations between process variables. However, there are also faults that difficult to be detected (fault 3, 9 and 15), because they are very small and have little influence to the process. According to Kano et al [2002a], all monitoring methods are unsuccessful in detecting fault 4, 5, 12 and 15, because these faults can be easily compensated by the control system. Therefore, the present paper mainly focuses on fault 3, 9, 15, 4, 5 and12.

For online fault detection, a data set obtained from an independent operation under normal condition is selected for the reference data set. In order to smooth the monitoring process, the size of the initial window is chosen as $w = 300$. According to the autocorrelation analysis, the modified dynamic step for each variable is chosen as shown in Table 2, where the cut-off value is chosen as 0.1. The weighted factor is chosen as $\lambda = 0.5$. Besides, 8 independent components and 18 principal components are selected for dynamic process monitoring methods. The results are summarized in Table 3.

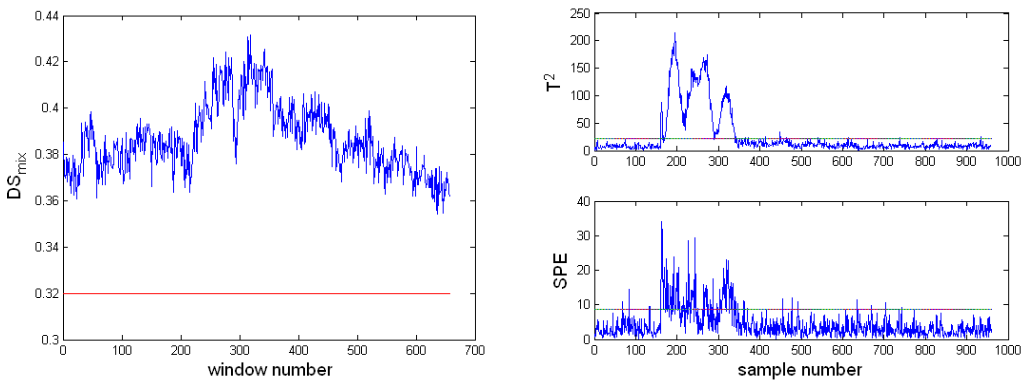
"DSF" refers to the dynamic similarity factor based method, and "MDSF" is the modified dynamic similarity factor based method. The maximum detection rate achieved for each fault is marked with a bold number. As shown in Table 3, MDSF outperforms other methods for most of the fault modes. Particularly, the monitoring performance of fault 3, 9 and 15 are greatly improved. Unfortunately, the performance of fault 4 and 15 become worse than DSF. However, compared to the DPCA and modified DPCA (MDPCA) methods, the monitoring performance of all the considerable faults have been greatly improved. Among all the monitoring mentioned methods, the MDSF method gets the best monitoring performance. Compared to the monitoring results of Kano, et al. [2002a], the proposed method not only can detect the small faults, but also can detect the latent faults, which are easily compensated by the control systems.

As examples of the latent and small faults, the monitoring results of fault 5 (latent) and fault 3 (small) are shown in Fig. 3 and Fig. 4. Due to the length of the paper, only the results of the new method and conventional PCA are presented. As demonstrated in Ge and Song [2007], fault 5 can be easily compensated by the control system thus the fault can not be detected after sample 350. They had improved the performance of this case by using ICA-PCA method. However, the information of correlation between variables in the present paper is not as sufficient as that in Ge and Song [2007], because we only choose 16 variables for process monitoring, compared to 33 variables in Ge and Song. [2007] Based on the used information of the present paper, the monitoring result of ICA-PCA method is similar to the conventional PCA method, which is not shown here. As shown in Fig. 3 (a), the fault is successfully detected by the proposed method. Fault 3 is considered as a small fault which has little influence to the process. The conventional PCA method can not detected the fault at all. However, because in this case the distribution of process data is changed, the monitoring performance is greatly improved, which is shown in Fig. 4 (a).

Next we consider the fault diagnosis ability of the new method upon these latent and small process faults. One or two responsible variables for the related fault are listed in table 4. Similar results can be found in Chiang, et al.[2000] except for fault 3, 9, and 15, which are not considered in their book. However, examining the cause of these three faults, the diagnosis results seem not very good. Hence, some improvement should be made to guarantee the reliability of fault diagnosis upon this method.

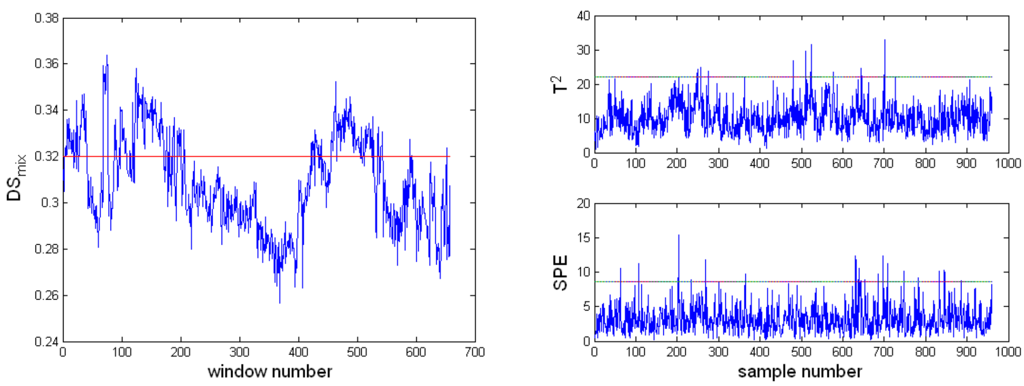
Table 1. Monitoring variables in the TE process

No.	Measured variables	No.	Measured variables
1	A feed	9	Product separator temperature
2	D feed	10	Product separator pressure
3	E feed	11	Product separator underflow
4	A and C feed	12	Stripper pressure
5	Recycle flow	13	Stripper temperature
6	Reactor feed rate	14	Stripper steam flow
7	Reactor temperature	15	Reactor cooling water outlet temperature
8	Purge rate	16	Separator cooling water outlet temperature



(a) (b)

Fig. 3. Monitoring results of fault 5, (a) new proposed method; (b) PCA



(a) (b)

Fig. 4. Monitoring results of fault 3, (a) new proposed method; (b) PCA

Table 2. Dynamic steps of monitoring variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Steps	1	0	0	0	0	0	0	1	1	3	0	3	3	3	0	0

Table 3. Monitoring results comparison

Fault modes	DSF	MDSF	DPCA(T^2)	DPCA(SPE)	MDPCA(T^2)	MDPCA(SPE)
3	20.0	**33.0**	1.0	2.3	2.1	1.7
4	**27.1**	17.8	1.2	1.3	1.4	2.2
5	89.4	**100**	26.0	26.8	27.0	28.1
9	35.8	**47.6**	2.1	2.0	3.8	2.3
12	64.6	**100**	96.7	98.1	98.2	98.8
15	**36.8**	32.2	3.5	1.7	3.8	3.5

Table 4. Fault diagnosis results of the considered faults

Faults	3	4	5	9	12	15
Responsible variables	6	4, 11	16, 10	7	15, 8	2, 3

6. CONCLUSIONS

In order to improve the monitoring performance of the latent and small faults for dynamic processes, a novel fault detection and diagnosis method has been proposed based on similarity factor and moving window. Compared to the conventional methods, the monitoring performance has been greatly improved, not only for the small faults, but also the latent faults of the process.

ACKNOWLEDEMENT

This work was financially supported by the National Natural Science Foundation of China. (Grant#60774067)

REFERENCES

Chiang LH., E.L. Russell, and R.D. Braatz. Fault detection and diagnosis in industrial systems. Springle-Verlag, London, 2001.

Chen Q., U. Kruger, and A.TY Leung. Regularised kernel density estimation for clustered process data. Control Engineering Practice. 12: 267-274, 2004.

Ge Z.Q., Z.H. Song. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors. Ind. Eng. Chem. Res. 46: 2054-2063, 2007.

Kano M., S. Hasebe, I. Hashimoto, and H. Ohno. Statistical process monitoring based on dissimilarity of process data. AIChE Journal, 48: 1231-1240, 2002a.

Kano M., K. Nagao, S.i Hasebe, I. Hashimoto, H. Ohno, R. Strauss, and B.R. Bakshi. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. Compu. & Chem. Eng. 26: 161-174, 2002b.

Krzanowski W.J.. Between-groups comparison of principal components. J. of Amer. Star. Assn. 74: 703-707, 1979.

Ku W., R.H. Storer, and C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. Chem. Intell. Lab. Syst., 30: 179-196, 1995.

Lyman P.R., C. Georgakist. Plant-wide Control of the Tennessee Eastman Problem. Compu. & Chem. Eng. 19: 321-331, 1995.

Singhai A., D.E. Seborg. Evaluation of a pattern matching method for the Tennessee Eastman challenge process. J. Process Control. 16: 601-613, 2006.

Zhao C.H., F.L. Wang, and M.X. Jia. Dissimilarity analysis based batch process monitoring using moving windows. AIChE Journal, 53: 1267-1277, 2007.