

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Learning based on kernel-PCA for abnormal event detection using filtering EWMA-ED.

José M. Bernal-de-Lázaro¹

Reference Center for Advanced Education, Cujae, Habana, Cuba.

Orestes. Llanes-Santiago²

Alberto. Prieto-Moreno³

Department of Automation and Computing, Cujae, Habana, Cuba.

Diego. Campos-Knupp⁴

Antonio. Silva-Neto⁵

Department of Mechanical Engineering and Energy, Polytechnic Institute, IPRJ-UERJ, RJ, Brasil

Abstract.

Multivariate statistical approaches have been widely applied to monitoring complex process, however incipient and small-magnitude faults may not be properly detected with the above techniques. In this paper, a learning approach based on kernel-PCA with filtering EWMA-ED is proposed to improve the detection of these types of faults. The proposal was tested on the Tennessee Eastman (TE) process where it is observed a significant decrease in the missing alarms, whereas the latency times are reduced.

Keywords. Fault Detection, kernel PCA, Small-magnitude faults, EWMA

1 Introduction

The abnormal event detection is the first crucial point of the fault diagnosis process [2]. With this purpose, many Multivariate Statistical Process Monitoring (MSPM) techniques as the Hotelling's T^2 statistic and the Squared Prediction Error (SPE) has been intensively studied [2, 4, 6]. However, in complex industrial process, some faults have a small effect on the monitored system and they can be hidden by the effect of disturbances and abrupt faults; causing lower performances in the fault monitoring process [5, 6]. In this paper, is proposed a learning approach based on kernel-PCA to improve the abnormal event detection associated with these fault types. The study is performed using the benchmark Tennessee Eastman (TE) process and all techniques are executed in Matlab®(R2015). The remaining of the paper is organized as follows. In Section 2, it is described the novel

¹jbernal@crea.cujae.edu.cu

²orestes@tesla.cujae.edu.cu

³albprieto@electrica.cujae.edu.cu

⁴diegoknupp@iprj.uerj.br

⁵ajsneto@iprj.uerj.br

EWMA with enhanced dynamic approach. The kernel PCA and the criteria employed for the adjustment of the kernel parameters are presented in Section 3. The used benchmark and the performance of fault detection systems are discussed in Section 4. Finally, based on the analysis of the results, the conclusions are given.

2 EWMA with Enhanced Dynamic Approach.

The Multivariate Statistical Process Monitoring (MSPM) strategies help to maintain the operations at industrial processes pointing out any anomalies in their behavior [2]. The Exponentially Weighted Moving Averages (EWMA) is a MSPM approach which has been used as filter of the SPE and Hotelling's statistics to detect small abnormal changes in the processes [5]. However, the combined use of this strategy with kernel methods, has not had high performances to detect incipient and small-magnitude faults.

A possible explanation for such low performance could be associated with an insufficient depth of EWMA's memory due to, in practice, achieving a well-adjusted memory in order to improve the Fault Detection Rates (FDR) without increasing the False Alarm Rates (FAR) is a difficult task. The EWMA with Enhanced Dynamic (EWMA-ED) herein developed, improves the detection of small-magnitude faults and it makes easier the setting of the memory parameter based on the dynamic for the classic SPE and Hotelling's T^2 statistics. The proposed filtering EWMA-ED can be calculated by:

$$\hat{\Upsilon}_t = \delta\gamma \sum_{j=1}^t (1-\gamma)^{t-j} \Upsilon_t + (1-\gamma)^t \Upsilon_0 \quad (1)$$

where Υ is the current value from an analyzed statistic (SPE or T^2 Hotelling) and $\hat{\Upsilon}_t$ is the corresponding estimated value, considering the behavior previously observable since the estimation Υ_0 . Note that, in Eq. (1), the proposed approach operates as a low pass filter where the weights $\sum_{j=1}^t (1-\gamma)^{t-j}$ geometrically decrease over time. However, γ is a constant ($0 < \gamma \leq 1$) that determines the depth of memory used by the EWMA approach and δ is an enhancer parameter, such that $0 < \delta \cdot \gamma \leq 1$. Consequentially, in this proposal, the enhancer factor (δ) on EWMA-ED, gives more importance to the most recent observation (i.e., the current value of the T^2 and SPE statistics) without ignoring the potential information contained in past samples. The EWMA scheme with enhanced dynamic smooths the behavior of the statistics reducing the difference between the weights in each sampling. As a result, the depth of memory is expanded and the influence of the present predicted value is more remarkable without losing sensitivity to small process shifts. The application of the EWMA-ED approach involves the use of conventional SPE and Hotelling's T^2 statistics and they thresholds, which can be calculated by:

$$T^2_{lim} = \frac{q(m^2 - 1)}{m(m - q)} F_{(q, m-q, \vartheta)} \quad (2)$$

where $F_{q, m-q, \vartheta}$ is the F -distribution with significance level ϑ and $(m - q)$ represents the freedom degree calculated by a number of retained components q and m samples. The

threshold for SPE is approximated by

$$\text{SPE}_{lim} = \theta_1 \left[\frac{h_0 c_a \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right]^{1/h_0}; \quad \theta_i = \sum_{j=a+1}^n \lambda_j^i; \quad h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (3)$$

3 Kernel Principal Component Analysis

The most disseminated feature extraction technique in the scientific literature is the Principal Component Analysis (PCA). This technique determines a set of transformation vectors sorted by the amount of explained variance that guarantees a reduction in the dimension which is optimal in terms to capture the variability in the information. However, the main disadvantage of the traditional PCA is its linear nature. [7]. For cases where exist nonlinear relationships the so-called kernel PCA, was introduced by [10]. KPCA solves the problems of nonlinearity using the basic idea of Cover’s theorem which states that a data structure of a nonlinear nature in an input space can be mapped to a higher dimensional space where it can be linearly separable [9]. In other words, KPCA does the PCA process in kernel space which introduces the advantage of high dimension mapping of original data using a kernel trick. According to [9], the steps to perform KPCA and reducing the dimension of the historical operating data can be described as illustrate the Figure 1.

Input: Historical data: $S_{N \times p} = \{\mathbf{X}_{m \times p}^1, \dots, \mathbf{X}_{m \times p}^c\}^T$, for c class;

- 1: $K_{ij} = K(x_i, x_j), \quad i, j = 1, \dots, N$;
- 2: $K - \frac{1}{N} \mathbf{j} \mathbf{j}' K - \frac{1}{N} K \mathbf{j} \mathbf{j}' + \frac{1}{N^2} (\mathbf{j}' K \mathbf{j}) \mathbf{j} \mathbf{j}'$;
- 3: $\alpha^j = \frac{1}{\sqrt{\lambda_j}} \mathbf{v}_j, \quad j = 1, \dots, N$;
- 4: $\tilde{x}_i = \left(\sum_{j=1}^N \alpha_j^j K(x_i, x_j) \right)_{j=1}^d$

Output: Transformed data: $\tilde{S} = \{\tilde{\mathbf{X}}_{m \times d}^1, \dots, \tilde{\mathbf{X}}_{m \times d}^c\}^T$

Figure 1: Pseudo-code of the algorithm kernel PCA.

3.1 Choosing a kernel.

In this paper, the RBF kernel was selected for its generality and results in multiple applications. The RBF kernel is defined by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right) \quad (4)$$

where the parameter σ is called bandwidth and indicates the degree of smoothness of the function [9]. The variation of this parameter must be done very carefully [1]. In this paper, the algorithm Differential Evolution (DE) is employed to simultaneously adjust the kernel parameter (σ) for KPCA, the depth of memory to the EWMA approach (γ) and the proposed enhancer parameter (δ). For this purpose, it was employed the DE/ $X^{\text{best}}/2$ bin scheme with a binomial distribution function. Additionally, given c mutually exclusive operating states (faults), different from the normal operating condition (NOC), the follow criterion was used as fitness function to adjust the RBF kernel:

$$\text{Fitness function} = 1 - \frac{1}{c} \sum_{i=1}^c \text{AUC}^i \quad (5)$$

The basic idea behind this criteria is to minimize the probability of error for the discrimination process corresponding to each pair (NOC, AOC^i) with $i = \{1, 2, \dots, c\}$, using the Area Under the Curve ROC (AUC) as a quality measure from each binary classification process. The Differential Evolution (DE) was executed using a population size $\text{NP} = 10$, a maximum iteration $\text{MaxIter} = 200$, a difference vector scale factor $F = 0.75$ and a crossover criterion $\text{CR} = 0.6$. Also, the search ranges for the estimated parameters was $\sigma \in [300, 1500]$, $\gamma \in [0.001, 1]$, $\delta \in [0.001, 100]$, respectively. The stopping criteria that were applied herein included the number of iterations and the value reached by the objective function.

4 Results and Comparative Analysis

In this section, the proposed approach is applied on the Tennessee Eastman (TE) process aiming to detect the simulated faults in this industrial benchmark.

4.1 Tennessee Eastman process.

The Tennessee Eastman (TE) process is widely used as a chemical plant benchmark to evaluate the performance of new control and monitoring strategies [1, 5, 6, 8, 11]. The process consists of five major units interconnected: a reactor, a condenser, a recycle compressor, a separator, and a stripper. The control objectives, suggested potential applications and features of the process simulation are described in more detail by [2] and [3]. The historical dataset of this benchmark, are generated during 48h with the inclusion of faults after 8 simulation hours. Each historical dataset contains a total of 52 variables (41 measured variables plus 11 manipulated variables) with a sampling time of 3 min and gaussian noise incorporated in all measurements. For the study performed in this paper, only the 33 variables available online are considered. They allow the analysis of the small-magnitude faults listed in Table 1. All datasets used in this paper can be downloaded from http://web.mit.edu/braatzgroup/TE_process.zip.

Table 1: Description of the analyzed faults in TE process.

Fault	Process variable	Type
3	D feed temperature	step
9	D feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
15	Condenser cooling water valve	Sticking

4.2 Results of the on-line application.

In order to monitor the faults of the TE process, the historical datasets should be pre-processing using KPCA and the confidence limit from SPE and Hotelling's T^2 statistics

has been previously established. In this paper, the number of principal components utilized by KPCA was 19 PCs, which explains 95.07% of the total process information. Consequently, the dimension of feature space in both cases was reduced to \mathbf{R}^{19} , according to the cutoff value of 0.001 for the eigenvalue and the estimated parameters $\sigma = 1276.23$, $\gamma = 0.006$ and $\delta = 3.05$. Once established the configuration of the tools, can be performed the on-line detection process. For this purpose, is recommended that the kernel PCA, the Hotelling's T^2 statistic and the proposed EWMA-ED approach are embedded on one multivariate detection scheme, which called kernel-MEWMA with reinforcing dynamic. Based on this, Figure 2 illustrates the multivariate detection scheme used in the present paper.

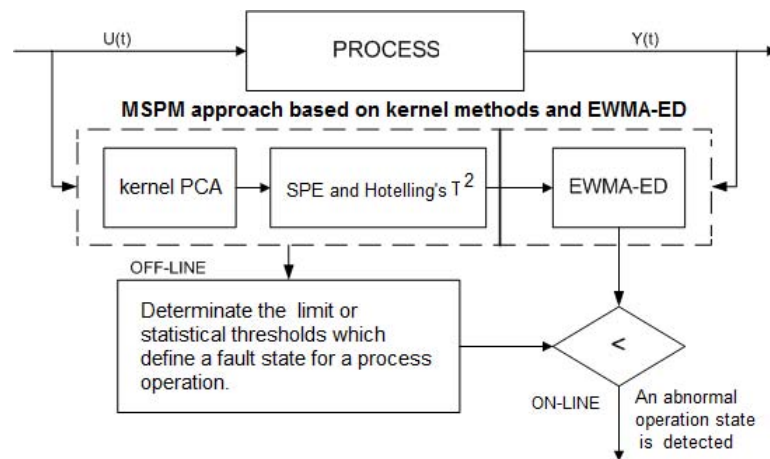


Figure 2: kernel-MEWMA with reinforcing dynamic

Using the above scheme, Figure 3 shows the monitoring charts obtained from Faults 3, 9 and 15, which are detected using the Hotelling statistic. Fault 3 is generated from one step in the D feed temperature, nevertheless the corresponding measures are quite similar to the normal data in terms of the mean and variance. As consequence, the conventional MSPM methods can not detected this fault appropriately. Figure 3(a) shows the graphical performance obtained by the scheme with EWMA-ED used to detect this fault. Figure 3(b) illustrates the performance of the T^2 statistic for Fault 9, which results from one random variation in the D feed temperature. As it can be seen from this monitoring chart, the proposed scheme detect the Fault 9 after the 116th sample. It means that the proposed approach generates 54 false alarms as a result of its high sensitivity to this change. The last case shown in Figure 3(c) presents the monitoring chart of Fault 15, which is a sticking in the condenser cooling water valve. Similar to Fault 3, the historical dataset of this fault has little difference with the normal data. According to [2] and [6], traditional statistics can hardly detect this fault. However, the T^2 chart of the detection scheme with EWMA-ED indicates the presence of abnormalities with a continuously detection of fault since the 161th sample.

In order to complement these graphical results, in Table 2 the monitoring performances (FAR and latency) of some representative small-magnitude faults of the TE process are tabulated. Using this information, in the same Table 2, a performance comparison between

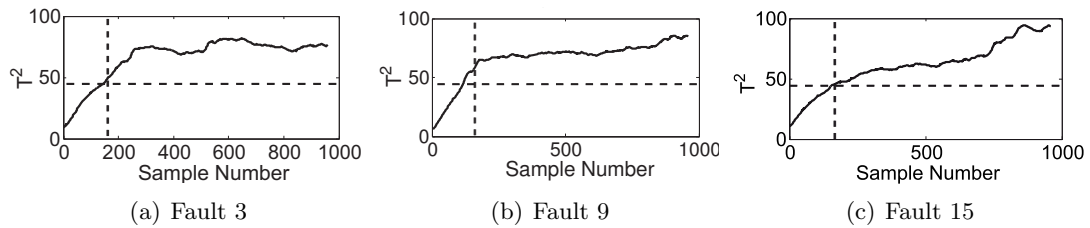


Figure 3: Hotelling’s T^2 charts obtained by the proposed approach to monitoring the faults 3, 9 and 15 from the Tennessee Eastman benchmark.

the approach herein proposed using the KPCA algorithm, and the Statistical-Local KPCA approach developed by [6], the Dynamic KPCA-based process monitoring applied by [4] (DKPCA) and the alternative Ensemble KPCA using by [8] (EKPCA) is additionally developed. The latency (L) is only analyzed for the proposed approach. It is an indicator that reflects the elapsed time before the fault will be continuously detected by the diagnostic system, considering the time in which the fault began (after 160th sample). In this paper, the continuous detection of a fault means that the diagnostic system emits five consecutive alarms pointing out that an abnormal operating condition is happening.

Table 2: Monitoring results and comparison with other approaches.

Faults	Proposed approach			SL-KPCA		DKPCA		EKPCA	
	T^2 (%)	SPE(%)	L(min)	T^2 (%)	SPE(%)	T^2 (%)	SPE(%)	T^2 (%)	SPE(%)
FDR 3	100.0	0.0	6.0	25.5	0.6	4.4	9.6	18.25	6.50
9	100.0	32.3	3.0	16.9	0.0	5.9	10.0	14.37	100.0
11	100.0	0.0	3.0	88.3	48.8	33.6	91.0	84.0	48.5
15	100.0	31.6	3.0	25.7	14.0	7.3	13.6	25.37	13.88

A detailed analysis of this comparison illustrates that, in terms of fault detection, the Hotelling’s T^2 have better performance indicator than the SPE statistic. Also, in general, the EWMA-ED based scheme have a better performance than the others approaches for all small-magnitude faults analyzed. As a result of this, may be said that the detection of Faults 3, 9 and 15 requires to consider the dynamic behavior of the process to improve all performance indicators, including the latency time.

5 Conclusions

In this paper, a novel multivariate statistical approach to detect abnormal events associate with small fault conditions in complex industrial processes was proposed. The proposed approach uses kernel PCA to do an effective capture of the nonlinear relationships among the process variables and the EWMA-ED approach in order to improve the monitoring of incipient and small-magnitude faults. Experiments have shown that the combination of the proposed approach with this kernel method helps to attain high performance rates in detection of small-magnitude faults. For future researches, may be interesting to investigate the use of others kernel methods and the application of different optimization techniques to tuning the kernel parameters.

Acknowledgments

The authors acknowledge the financial support provided by FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, from Brazil, and CUJAE, Instituto Superior Politécnico José A. Echeverría from Cuba.

References

- [1] J. M. Bernal-de-Lázaro, A. Prieto-Moreno, O. Llanes-Santiago and A. Silva-Neto. Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers & Industrial Eng.*, 87:140–149, 2015.
- [2] L. H. Chiang, R. D. Braatz and E. L. Russell. *Fault detection and diagnosis in industrial systems*. Springer, 2001.
- [3] J. J. Downs and E. F. Vogel. A plant-wide industrial process control problem. *Computers & Chemical Eng*, 17(3):245–255, 1993.
- [4] J. Fan and Y. Wang. Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Information Sciences*, 259:369–379, 2014.
- [5] J. Fan, S. J. Qin and Y. Wang. Online monitoring of nonlinear multivariate industrial processes using filtering KICA-PCA. *Control Eng. Practice*, 22:205–216, 2014.
- [6] Z. Ge, C. Yang and Z. Song. Improved kernel PCA-based monitoring approach for nonlinear processes. *Chemical Eng. Science*, 64(9):2245–2255, 2009.
- [7] M. Jia, H. Xu, X. Liu and N. Wang. The optimization of the kind and parameters of kernel function in KPCA for process monitoring. *Computers & Chemical Engineering*, 46:94–104, 2012.
- [8] N. Li and Y. Yang. Ensemble kernel principal component analysis for improved nonlinear process monitoring. *Industrial & Eng. Chemistry Research*, 54(1):318–329, 2014.
- [9] J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.
- [10] B. Schölkopf, A. Smola and K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10: 1299–1319, 1998.
- [11] Y. Zhang. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chemical Eng. Science*, 64(5):801–811, 2009.