Abnormal Process State Detection by Cluster Center Point Monitoring in BWR Nuclear Power Plant

J. Talonen, and M. Sirola
Department of Information and Computer Science, Helsinki University of Technology, Finland

Abstract—*This paper proposes a new method to detect abnormal process state. The method is based on cluster center point monitoring in time and is demonstrated in its application to data from Olkiluoto nuclear power plant. Typical statistical features are extracted, mapped to n-dimensional space, and clustered online for every time step. The process signals in the constant time window are classified into two clusters by the K-means method. By monitoring features of the process signals, in addition to signal trends and alarm lists, the operator gains a tool that helps in early detection of the pre-stages of a process fault. By using cluster center point time series monitoring, faults in the process can be seen by at first glance or automatically by notification in the alarm list. This provides a definite advantage to any operating personnel and ultimately improves safety at the nuclear power plant.*

Keywords: nuclear industry, abnormal process state detection, high dimensional data, feature extraction, classification.

1. Introduction

The goal of the process state detection method presented in this paper is to detect abnormalities in Olkiluoto boiling water reactor (BWR) type nuclear power plant (NPP) in Finland. At Olkiluoto, thousands of signals are measured and monitored. Because of the high dimensionality of the system, manual selection becomes arduous. When a large numbers of process signals exist, subsets of relevant variables are automatically selected for modeling [1], [2]. This paper, however, does not focus on the variable selection phase. It is assumed that variables can be selected from certain area of the plant or from all around the NPP. In other words, variable selection depends on the need to improve given monitoring in a certain area. The main emphasis of this paper is placed on the cluster center point movement monitoring of those variables.

An earlier study was conducted to investigate classification by principal component analysis (PCA) [2]. The sizes of the groups are same, and each object can be assigned to many groups. In this paper, signals are classified in two categories: *slow* (steady, inactive) and *fast* (quick, noisy). Therefore data is classified by the K-means method into two clusters for every time step of a constant frame size. The sizes of the groups are different, and each object is assigned only to one group. During a normal operation state,

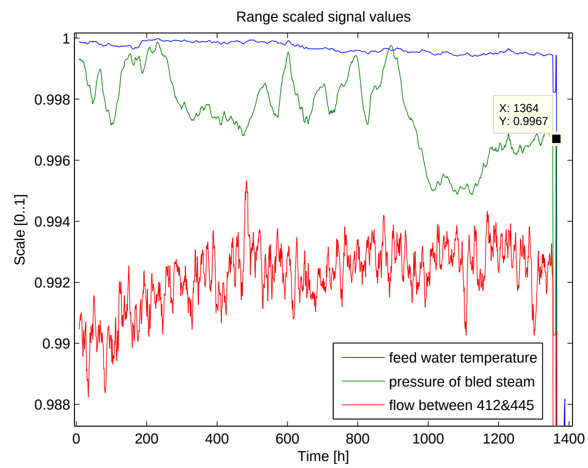

Fig. 1: The time series for three preprocessed process signals. The zero value is the global minimum (stored minimum value of the signal in the database) for the signal and one is the maximum. Most of the variables are near the global maximum value in the normal operation state but this naturally depends on the type of the variable.

most of the signals are classified as *slow*. In an abnormal process state process signals vary and are classified as *fast*. By monitoring features of the process signals the pre-stages of the process fault can be detected.

Within the Olkiluoto control room, there is an overload of alarms and notification which make it difficult for the operator to make discerning decision. Some sort of alarm sanitation is required [3]. The need for alarm handling is reduced if there are meaningful and clear statistics derived from process data [4], [5], [6]. For example: Hotelling's T^2 statistics can be used to detect faults for multivariate process data. This methos is actually compared to our method in the section *simulation results*. Other monitoring systems based on noise analysis also already exist [7]. If a fault or its pre-stage is detected, large-scale systems like the Olkiluoto NPP can be improved significantly. In our newest research and in this paper, real data from reactor unit 2 of Olkiluoto NPP is used [8]. In 2007, more than 300 averaged signals were stored, every hour, over a two months period. During this time, an abnormal process state in the turbine section of the NPP was captured in the recorded data.

2. Description of used methods

The traditional way in industrial plants to monitor time series is with Shewhart charts and limit value checking [6]. Monitored signals have lower control limit (LCL) and upper control limit (UCL). With this method it is difficult to find the correct target value and limit values for each signal. One reason for this is that the industrial process generates many different types of signals. For example, Bergquist introduced 14 different signal classes [9]. These classes are periodic, slowly varying, multiple steady state, and containing outliers. Three different signals from Olkiluoto NPP are shown in the Figure 1.

Features from range scaled signals are measured for each time step. Scaling of the signals is important because without it, signal values or features cannot be compared between other signals [2], [10]. The results cannot be reliably clustered without preprocessing the signals, so the effect of white noise is eliminated by moving average (MA)

$$\overline{x}_t = \frac{1}{N_m} \sum_{k=0}^{N_m-1} x_{t-k}, \tag{1}$$

where x_t is a scaled measurement value and N_m is a *frame size* of the moving average. First, the $N_m - 1$ data points are removed from the beginning. The undesirable start effect is erased, which is acceptable because the data is stored continuously and in actuality no data is wasted.

The first feature in this application is the *absolute difference,* which is used to measure the rate of change

$$d_t = \frac{|\overline{x}_t - \overline{x}_{t-N_d}|}{N_d}, \tag{2}$$

where x_t is a preprocessed (scaled and averaged) measurement value and N_d is a *frame size* of the difference. Difference is high pass filter and it extracts changes in the signal. The Second feature is the moving standard deviation (MSDV). It is a common measure of statistical dispersion and it measures how widely the values are spread in time. If the data points are far from the mean, then the MSDV values are large. If all the data values are equal, then the MSDV value at the current time is zero. MSDV is derived from the *sample variance*

$$MSDV_t = \sqrt{\frac{1}{N_s - 1} \sum_{k=0}^{N_s-1} (x_{t-k} - \overline{x}_t)^2}, \tag{3}$$

where N_s is the *frame size* for MSDV, x_t is a scaled sample value and $\overline{x}_t$ is the moving average [11].

In this paper only two statistical features are introduced, but there could be more such as skewness (measure of symmetry of a distribution) and kurtosis (measure of the peakedness or flatness of a distribution when compared with a normal distribution). Selected features depend on the goal of the classification.

K-means method is an unsupervised learning algorithm, which classifies a given data set through a certain number of k clusters [12]. The initial placement of the centroids are randomly defined for every time step, one for each cluster. Each object is assigned to the group that is closest to the centroid. When all objects have been assigned, the positions of the k centroids are recalculated. These steps are repeated until the centroids no longer move. The optimal solution is to minimize the cost-function

$$J = \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2, \tag{4}$$

where there are k clusters S_i, $i = 1, 2, \ldots, k$ and μ_i is the centroid of all the points $x_j \in S_i$. With these features it is clear that the center point near the axes origin relates to the *slow* signals. The absolute value of the rate of change and the MSDV are small compared to the *fast* signals.

The operators in the control room are already overloaded with monitoring work, so a simple index, *unsteadiness*, is introduced. The idea is to monitor the cluster center point coordinates of the *slow* signals. It is more important to concentrate on the *slow* signals because there are remarkably large changes in some measurements in their normal operating state. These signals temporarily have high MSDV and difference values. These rapid but normal changes in process signal values increase the center point coordinates of *fast* signals. Examples of such events in NPP are: control flow from one pipeline to another, watering, and rapid changes in flow in a feed filter. The *unsteadiness* limit can be adjusted to produce an automatic alarm. The MA of the cluster center point of the *fast* signals is measured and the current alarm limit its minimum value.

Our method is compared to the Hotelling's T^2 statistics which is a measure of the variation within the PCA model.

$$T^2 = (\mathbf{H} - \overline{\mathbf{H}})^T \mathbf{S}^{-1} (\mathbf{H} - \overline{\mathbf{H}}), \tag{5}$$

where $\mathbf{H}$ is the score matrix, $\overline{\mathbf{H}}$ and $\mathbf{S}$ are the common estimators for the mean vector and covariance matrix obtained from the scores [13]. The scores $\mathbf{H}$ are the preprocessed data mapped into the new coordinate system defined by the principal components.

Process abnormality is detected with the help of Hotelling's T^2, which defines the normal operating area corresponding to 95% confidence. The upper control limit (UCL) of the multivariate Hotelling's T^2 statistics can be defined

$$T^2_{UCL} = \frac{(n-1)(n+1)k}{n(n-k)} F_\alpha(k, n-k), \tag{6}$$

where $F_\alpha(k, n-k)$ is the upper critical point of the F-distribution with k and $n-k$ degrees of freedom. In practice k is the amount of selected variables and n is the number of measurements. [6]

3. Simulation Results

All variables are range scaled by the database minimum and maximum values because it provides better classification results. In our research the database was constructed offline. It was aggregated by 40 design based and abnormal stored data sets such as watering and scram situations. These data sets are provided by Teollisuuden Voima Oy. The frame size used for the MA, MSDV, and *absolute value of the rate of change* was eight hours.

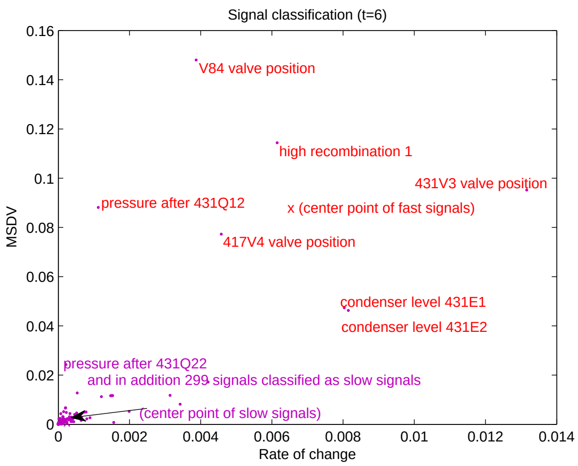

Fig. 2: An example of the clustering result at $t = 6$. Variables located near the axes origin are classified as *slow* signals. Variables near the other cluster center point (X) are classified as *fast* signals.

Signal measurements were selected all around the NPP, because it was decided that common safety improvements were a priority. Features are measured for 307 variables at each time step. The *unsteadiness* in this case monitors the general process state in the NPP. Feature values of each variable, the classification result, and the center points can be illustrated and updated in selected time period (every hour or minute), see Figure 2. This can be useful for expert users but not for operators in the control room.

In Figure 3 the deviation (the center point of MSDVs) for the *slow* and the *fast* signals is shown. The automatic alarm is based on these values. Deviation of the *slow* signals increased and NPP is not stable after $t = 1361$. Because of the limit was exceeded, the *unsteadiness* notification is shown in the display of alarms at the control room. In this visualization, high MSDV values for the *fast* signal can be seen at $t = 961$. It is a normal operation state and the high measurement values are caused by rapid changes in feed filters flows and temperatures. Few variables change the MSDVs and they only have a minor effect to the center point of *slow* signals.

Figures 4 and 5 show time series which can be displayed in the control room. These are actually not mandatory

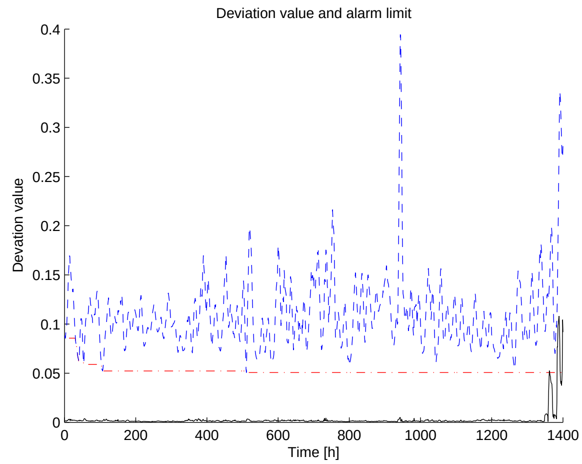

Fig. 3: Deviation values for the *slow* signals (solid line) and the alarm limit (dash dotted line). It is the minimum value of the moving average of deviation for the *fast* signals (dashed line) until current time.

because a notification system using the *unsteadiness* index works without operator monitoring. Of course important information is aggregated to the limit line as in the case of the center point coordinates of the *fast* signals. After the alarm notification, the limit can be re-settled. In this case the notification of *unsteadiness* is given at $t = 1359$ because of the high absolute difference values and at $t = 1361$ because of the high deviation values.

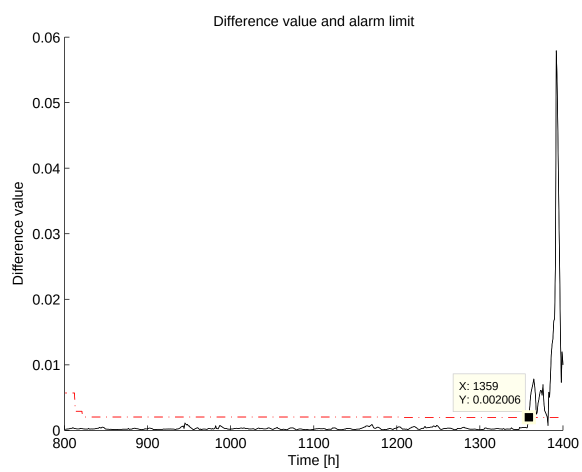

Fig. 4: The absolute difference value of the *slow* signals (solid line) and the minimum value of the *fast* signals cluster center points (dash dotted line).

The simulation results are satisfying, because most of the signals get normal values until $t = 1364$, see Figure 1. The method presented here detects a process fault three hours

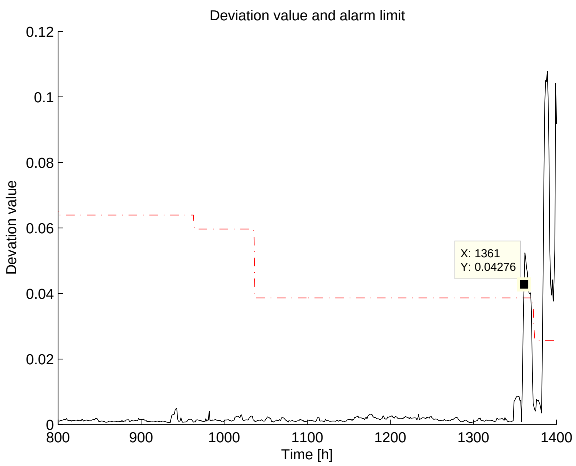

Fig. 5: MSDV of *slow* signals (solid line) and the minimum value of the *fast* signals cluster center points (dash dotted line).

earlier than Hotelling's T^2 statistics, see Figure 6. There are some false alarms: at the time of rapid changes in feed filters flows and temperatures, $t = 961$, and when a gas flow was controlled from one pipeline to another at $t = 1175$.

PCA was performed offline all time series data with same preprocessing parameters. Complicated matrix inversions are usually a problem in multivariate methods. Therefore an online implementation of PCA would be a problem especially when the size of moving time window is small. Because of the feature extraction and data mapping into two dimensional space, these problems do not exist in method presented in this paper.

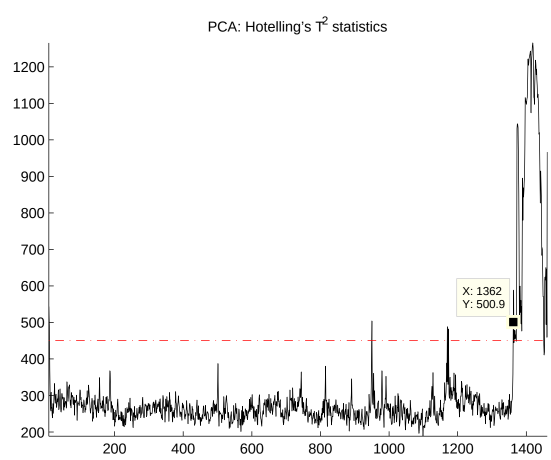

Fig. 6: Hotelling's T^2 statistics (solid line) and the upper control limit (UCL) as an alarm limit (dash dotted line).

4. Conclusion

The method presented in this paper can be used for pre-stage detection of process faults through the introduction of an unsteadiness index. The implementation of this method requires little investment in terms of increased process complexity while offering a solid advantage in terms of pre-stage fault detection capability. Through early detection, NPP will ultimately enjoy reduced maintenance costs, decreased downtimes, and increased safety.

One possible limitation for the method presented here is that the NPP has rather strict rules for the automation system changes. Also computation time can be a problem if data is analyzed too frequently. Future research will concentrate on developing methods for maintenance personnel. There is also a demand for the development of calibration monitoring technology because more than 90 percent of current calibration efforts are unnecessary. Calibration can also be harmful to reactor equipment and even to plant safety [14].

References

[1] S. Laine. *Using visualization, variable selection and feature extraction to learn from industrial data*. PhD thesis, Helsinki University of Technology, 2003.

[2] J. Talonen, M. Sirola, and J. Parviainen. Leakage Detection by Adaptive Process Modeling. In R. Stahlbock, S.F. Crone, and S. Lessmann, editors, *Proceedings or The 2008 International Conference on Data Mining (DMIN 2008)*, volume I, II, July 2008.

[3] J.E. Larsson. Simple Methods for Alarm Sanitation. In *Proceedings of the IFAC Symposium on Artificial Intelligence in Real-Time Control, Budapest*, 2000.

[4] J. Talonen. Fault Detection by Adaptive Process Modeling for Nuclear Power Plant. Master's thesis, Helsinki University of Technology, 2007.

[5] J. Chen and K.C. Liu. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science*, 57(1):63–75, 2002.

[6] L.H. Chiang, E. Russell, and R.D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2001.

[7] J. Ortiz-Villafuerte, R. Castillo-Durán, G. Alonso, and G. Calleros-Micheland. BWR online monitoring system based on noise analysis. *Nuclear Engineering and Design*, 236(22):2394–2404, 2006.

[8] M. Sirola, J. Talonen, and G. Lampi. SOM based methods in early fault detection of nuclear industry. In *Proceedings of the 17th European Symposium On Artificial Neural Networks ESANN'09*, April 2009.

[9] T. Bergquist, J. Ahnlund, and JE Larsson. Alarm reduction in industrial process control. In *Emerging Technologies and Factory Automation, 2003. Proceedings. ETFA'03. IEEE Conference*, volume 2, 2003.

[10] D. Pyle. *Data Preparation for Data Mining*. Morgan Kaufmann, 1999.

[11] J.F. Hair, R.E. Anderson, R.L. Tatham, and W.C. Black. *Multivariate Data Analysis*. Prentice Hall, 5th edition, 1998.

[12] S. Theodoridis. *Pattern Recognition*. Academic Press, 2003.

[13] L.I. Tong, C.H. Wang, and C.L. Huang. Monitoring defects in IC fabrication using a Hotelling T2 control chart. *Semiconductor Manufacturing, IEEE Transactions on*, 18(1):140–147, 2005.

[14] H.M. Hashemian. History of On-Line Calibration Monitoring Developments in Nuclear Power Plants. 27th-29th of September 2004. Halden.