ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Feature selection on process fault detection and visualization

Tuomas Alhonnoro and Miki Sirola

Helsinki University of Technology - Department of Computer and Information Science
P.O.Box 5400, 02015 TKK, Finland

Abstract. Feature subset selection has become an essential part in data mining applications. In this article, feature subset selection is integrated into real time process fault detection. Various methods based on both dependency measures and cluster separability measures are discussed. An intuitive tool for process visualization is introduced. Experiments on nuclear power plant simulator data are carried out to assess the effectiveness and performance of the methods. Early detection of failures, which is one important goal in the project, is achieved with help of visualizations developed in this work. In a leak scenario an illustrative example was produced.

1 Introduction

Methodology governing modern process fault detection and diagnosis can be divided into either quantitative or qualitative model-based methods and process history based methods [1]. In this paper, the latter is of concern, as a new intuitive Fault Indication SubSystem (FISS) is developed. The main focus is kept on Nuclear Power Plant (NPP) operation, although the FISS can be adapted into any industrial process.

Based on binary threshold values, the current power plant monitoring contains little information on the level the process is propagating. While additional information would lie on the graphs, the overall picture is merged among hundreds of different variables. These two drawbacks can be improved by introducing a new continuous sensor and filtering the number of concurrent variables depicted on control room. The latter utilizes Feature Subset Selection (FSS) filtering [2], the former is built on Self-Organizing Maps (SOM) [3]. Demand for renewal of monitoring is immense, as large-screen monitors are taking place in control rooms.

This work was carried out within a larger entity, called the NoTeS project (Nonlinear Temporal and Spatial forecasting: modeling and uncertainty analysis [4]). NoTeS project is developing a generic toolset for spatio-temporal forecasting and forecast uncertainty analysis through analyzing five widely different test cases, among which one, namely supporting operational decisions at NPP, covers this topic. Experimental part and testing of the methods were carried out with training simulator data kindly provided by the industrial partner Teollisuuden Voima Oy (Olkiluoto NPP).

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

2 Subset Selection: methodology

Feature Subset Selection is a preprocessing step that aims for a smaller set of representative features without losing the essential characteristics of the original data [5]. The main benefits of successful subset selection reads as 1) dimension reduction for less complexed computation, 2)removal of irrelevant and redundant variables for more compact representation, 3) better classification results with more generalizable models, 4) better understanding of essential characteristics of the underlying process and 5) compact visualization.

Traditional FSS methods are divided in filters and wrappers [2]. In this section, the very principles in feature selection filters are discussed.

2.1 Preprocessing

To apply statistics into a multivariate time-series data (MTS), the time-series is divided into a series of frames of equal length, which are then assumed quasistationary. Moreover, for statistics to perform correctly, every feature is normalized into zero mean and unit variance, thus $\mathcal{D}(\mu, \sigma^2) \rightarrow \mathcal{D}(0, 1)$. For anomaly detection purposes, normalization can be calculated with respect to regular conditions only.

2.2 Linear dependency

In linear sense, a set of features are considered the same if they *correlate*.

Correlation based feature selection can be built on distance matrix [6] $\mathbf{D} = \{D_{ij} | D_{ij} = 1 - |\rho_{ij}|\}$, where ρ_{ij} is the correlation coefficient between two random variables. Among every cluster, achieved by e.g. K-means algorithm [7], the feature closest to the cluster centroid is chosen as a representative feature.

2.3 Cluster Separability

While the former method seeks for the most dissimilar set of variables, the information theoretic approach maximizes the cluster separability based on assumption that at least two different clusters exist. The cluster separability can be defined as mutual information between two probability distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ and it is usually referred as *Kullback-Leibner divergence*, which is the entropy of p relative to q or

$$D(p||q) = \int_{-\infty}^{\infty} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}.$$

Similarly one can write $D(q||p)$. While KL-divergence is non symmetric, a new metric distance, say *KL-distance* is defined as $d_{pq} = D(q||p) + D(p||q)$.

The feature subset selection algorithms built on information theoretic dissimilarity are usually based on exhaustive search. These include e.g. conventional »greedy» or sequential algorithms (SFS, SBS and their variants) and Branch and

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Bound algorithm [8]. Alternatively, an intuitive technique, *scalar feature selection* [7] maximizes the information theoretic separability of individual variable rectified by it's correlations with already chosen variables.

2.4 Self-Organizing Maps

Self-Organizing Map (SOM) is an unsupervised neural network algorithm, which has its advantages in visualizing high-dimensional data. The SOM consists of neurons (*units*) organized on a regular two-dimensional grid. Each unit contains a model vector $\mathbf{m_i}$, that are updated according to equation

$$\mathbf{m_i}(t+1) = \mathbf{m_i}(t) + h_{ci}\left[\mathbf{x}(t) - \mathbf{m_i}(t)\right],$$

where $\mathbf{x}$ is an input data vector, $t = 0, 1, 2, \ldots$ is the number of learning step and h_{ci} is the neighborhood function, that controls how each unit is updated around the best matching unit. Thus SOM algorithm organizes the model vectors so that neighboring units in the grid represent similar data.

An ordered SOM grid is visualized in the U-matrix form (refer to Figure 1), which consist of hexagonals that represent grid unit densities as well as nearest neigbor distances by a certain colors.

SOM dissimilarity matrix $\mathbf{D}$ can be used for clustering as is [9], but more powerful is forcing the input data into a shrunked grid, where every grid represents one potential cluster.

3 Anomaly monitoring

The feature selection phase uses the first two methods for 1) removing redundancy and 2) weighting variables according to their importance. The assumption is that adding a subset selection filter on the process data improves the sensitivity of detection.

The second task is to evaluate the filtered data. In this work, process anomaly is detected by comparison of the current state with the so-called normal reference state, which is a sequence of normal operational data and thus the only history needed for detection. The time development is visualized on a special SOM U-matrix (see the next section). Subset selection can catch operator's attention on the most dubious variables. The decision and the actual diagnosis must be done by the operator itself on the basis of subset selection and expert knowledge.

3.1 SOM visualization: FISS

The implemented SOM visualization (see Figure 1) works as follows.

- Initialize the SOM U-matrix so, that the color of the *normal reference state* is fitted into a specified *normal color range*. This color range explains normal variation in process.

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

- Then, for every time step, draw U-matrix for the combination of the current data frame and the *normal reference frame.* If the process state develops into separate clusters, the distances between the neighboring units increase, thus the color exceeds the *normal color range.* An intuitive *full color range* is a colormap that runs through blue-green-yellow-red. A change from green to red is naturally associated with problems.

The implemented SOM U-matrix visualization is clear and natural indicator of possible fault. The interpretation of U-matrix is best described in Figure 1, which shows few moments of evolving leakage in the main circulation pump P1. The leftmost picture shows a normal event with no separation visible, thus the color is equal to *normal color range.* On the middle, there is an early detection of the leakage four minutes after the fault started evolving. According to tick marks, the process states are fully separated, but the difference between the frames is relatively low, although clearly visible with light orange colored border. Finally, the rightmost picture shows a moment just after the scram, where most primary variables are shut down, and thus the state separation is maximized.

Experiments on the data showed, that SOM U-matrix distances exhibit a pretty low variation regardless of the clusterability. That is, inspecting the color chances of the U-matrix would hardly give a rough picture of possible state separation. This flaw was outcome by nonlinear transformation of the U-matrix elements, thus

$$U_{ij} \rightarrow \exp(U_{ij}).$$

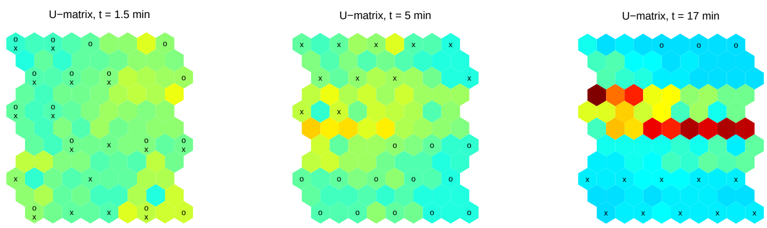

Figure 1: Scenario 1: 0-10% leakage in the main circulation pump P1. SOM-visualization of various moments. *Normal reference state* and current state are marked with »o» and »x» respectively.

4 Experimental results and discussion

4.1 Datasets

The dataset consists of four scenarios simulated at the operator's training centre in Olkiluoto NPP [10]. In addition, some synthetic data was constructed on the basis of simulated dataset.

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Table 1: Scenario 1: 0-10% leakage on main circulation pump P1. The most important events. Compare with the FISS-visualization in Figure 1.

time (min:s)	Event
1:00	The fault started evolving
5:48	Controlled area floor drain sensor triggered
9:34	Rotation difference in pump P1 detected
13:49	Reactor scram triggered by the leakage control

Each scenario simulates a propagation of a specified malfunction model, which includes e.g. leakage in main circulation pump or pressurizer failure. The scenarios simulate the actual behavior of automatic process control, thus no operational decisions were done and all the scenarios ended in emergency shut-down. Each scenario consists of 78 process variables recorded at each sampling instant. In addition the respective caution and alarm signals (Table 1) were printed for further use.

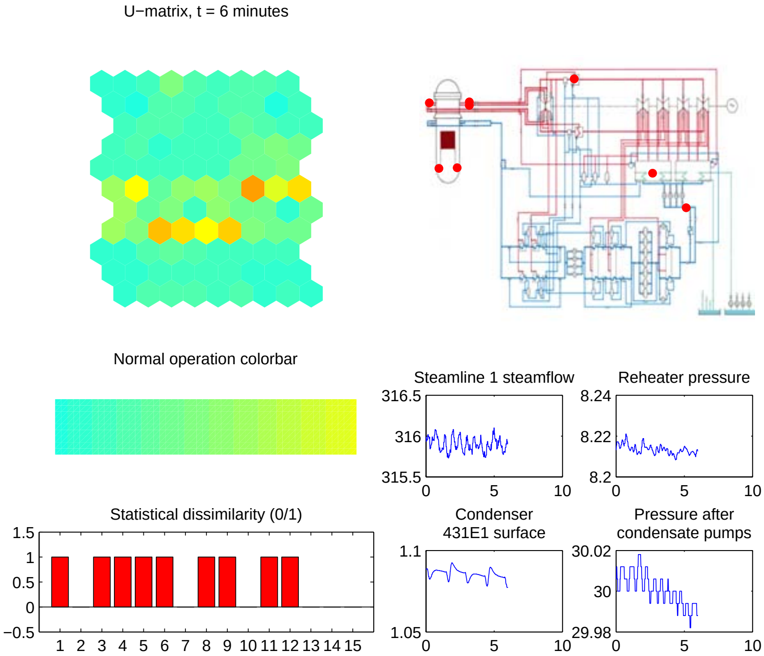

Figure 2: Control room visualization, Man-machine interface (MMI): FISS combined with statistical Kolmogorov-Smirnov test (KS-test), process flow diagram and selected process variable graphs. A red bar on a given variable indicates negative KS-test i.e. dissimilarity between two probability distributions.

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

4.2 Results and conclusion

Various experiments were carried out in MATLAB environment.

Subset selection algorithms proved useful in various situations. Removing the irrelevant variables clearly improved the separability of two different states. This, however, was not fitted into an objective comparison chart as the research topic itself is highly subjective – no two fault scenarios are comparable.

The greatest advantage of subset selection is connected to situations governed by small, local changes in the process. These include e.g. sudden vibrations in a pump, or local coolant losses. To emphasize, these small deviations are the most important for early detection. The special man-machine interface shown in the Figure 2 then serves as an auxiliary component for fault detection and isolation.

On the other hand, most process variables are highly dependent of each other, that is, for example, water discharging from the main circulation pump immediately affects surface levels and pressures all around the primary circuit. In such severe situations, subset selection hardly has any contribution on fault isolation. The FISS, however, can be used to track the evaluation of the process.

Finally, it should be pointed out, that selecting the FSS-method is a trade-off between high state separability (indexing for state separability) and process diversity (clustering for dissimilarity). By now, the best answer does not exist.

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S.N. Kavuri. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3):293–311, 2003.

[2] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(7-8):1157–1182, 2003.

[3] T. Kohonen. *Self-Organizing Maps.* Springer, Berlin, 3rd edition, 2001.

[4] R. Ritala, E. Alhoniemi, T. Kauranne, K. Konkarikoski, A. Lendasse, and M. Sirola. Nonlinear temporal and spatial forecasting: modeling and uncertainty analysis (NoTeS) - MASIT20. *MASI Technology Programme 2005 - 2009 Yearbook 2007*, 207:10, 2007.

[5] T. Alhonnoro. Feature subset selection for process fault detection. Special assignment, Helsinki University of Technology, 2007.

[6] P. Mitra, C.A. Murthy, and S.K. Pal. Unsupervised feature selection using feature similarity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):301–312, 2002.

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

[8] M. Kudo and J. Sklansky. Comparison of algorithms that select features for pattern classi ers. *Pattern Recognition*, 33(1), 2000.

[9] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas. SOM Toolbox for Matlab 5, 2000. SOM Toolbox available for download in `http://www.cis.hut.fi/projects/somtoolbox`, retrieved at August 2007.

[10] Anonymous. *Olkiluoto, Ydinvoimalaitosyksiköt Olkiluoto 1 ja Olkiluoto 2.* Teollisuuden Voima Oy. `http://www.tvo.fi/`, retrieved at July 2007.