Chapter 16

Data-based analysis of industrial processes

Olli Simula, Miki Sirola, Jaakko Hollmén, Ignacío Diaz, Sampsa Laine, Esa Alhoniemi, Johan Himberg, Jussi Ahola, Jukka Parviainen, Timo Similä, Jarkko Tikka

16.1 Extraction and identification of operational states of a paper machine

An underlying assumption in most data analysis, monitoring and diagnostics methods is that the process considered is in a stationary state. However, in practice this assumption is often ignored due to lack of tools for the analysis of operational states. Ignoring the stationarity assumption results in poorer performance of the monitoring and diagnostic methods. In paper production, there exist many operational states that affect the performance of the process, yet are not clearly distinguishable from measured data.

When a monitoring or diagnostics system for improved fault detection is implemented, a problem frequently arises: are the data used as reference for future process operation collected from a stationary state and is this state achievable repeatedly? Also, the stability of the process – and thus the effect of a same kind of disturbance – may greatly vary depending on various circumstances, which are not connected to the origins of the disturbance. Thus, another question concerning the states of a process is: if the same product can be produced in various ways, which is the most profitable?

In order to study and manage different process states, a cluster analysis tool was developed. It consists of all the modules that are required in the cluster analysis: preprocessing, clustering and interpretation parts. The whole framework with a description of the tool is reported in [1]. A method for interpretation and comparison of clusters – that are in this context seen as operational states of a process – was also reported separately in [2]. Some visualizations of one process state are shown in Figure 16.1.

References

[1] M. Federley, E. Alhoniemi, M. Laitila, M. Suojärvi, and R. Ritala. State management for process monitoring, diagnostics and optimization. *Pulp & Paper Canada* (to appear).

[2] E. Alhoniemi and O. Simula. Interpretation and Comparison of Multidimensional Data Partitions. In Michel Verleysen, editor, Proceedings of the 9th European Symposium on Artificial Neural Networks, pages 277–282. D-Facto, 2001.

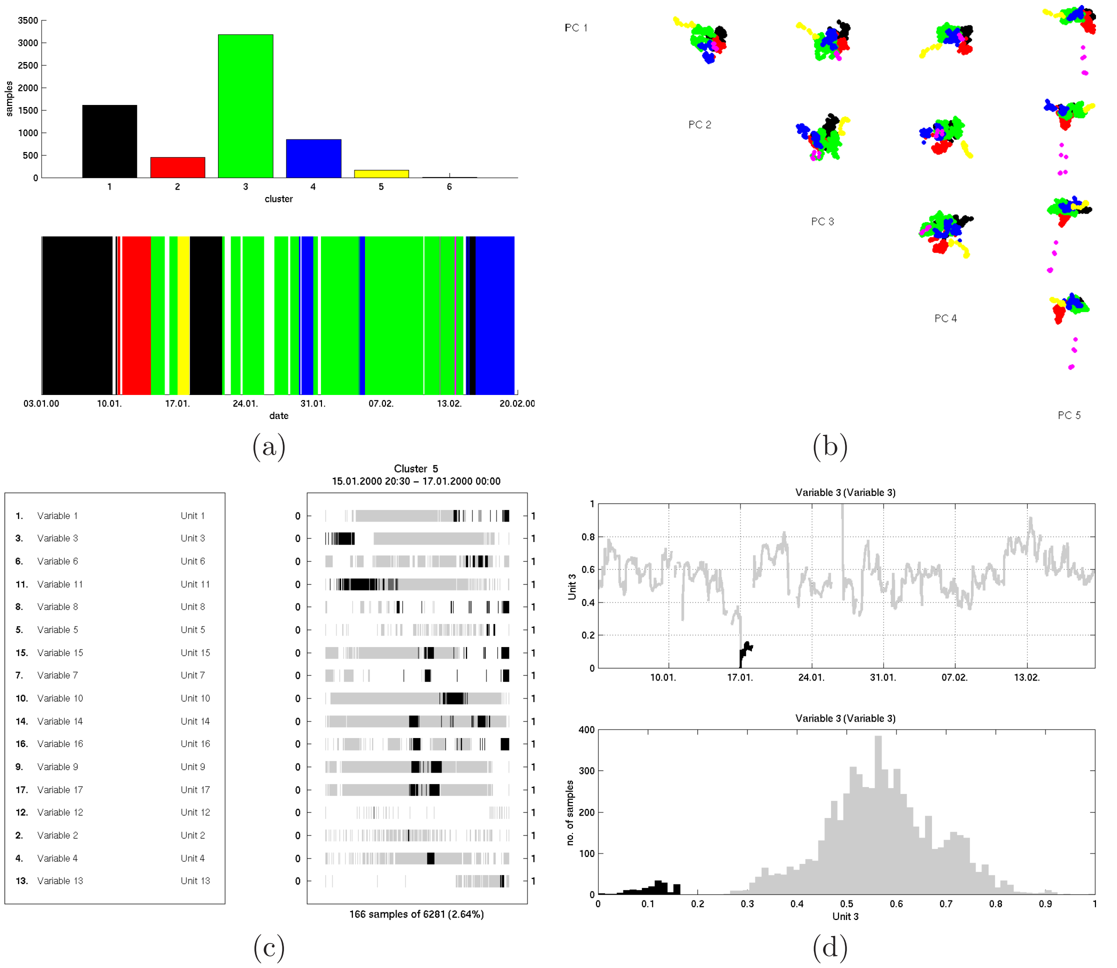

Figure 16.1: Different visualizations produced by the clustering tool. (a) Sizes of six extracted clusters and their location in time, (b) PCA projection of the clusters, (c) a "thumbnail" display of cluster 5 (which are marked by yellow color in (a) and (b)), and (d) a time series and histogram display of one variable (number 3) that is important in characterization of cluster 5.

16.2 Variable selection, feature extraction and visualization in data based process study

Consider yourself as a process engineer with a task of solving a recurring process control problem, for example, the disturbances frequently caused by the changes in the process feed type. The natural starting point is to collect information by interviewing process operators, discussing with colleagues and gathering process data. Then you would return to study the acquired information to define the problem, to define a solution, and to create an application to support process control. This project studies how the tools of variable selection, feature extraction and data visualization can support this task.

In this project, a full-scale mineralogical and a pilot scale pharmaceutical process were considered. In [1] manual variable selection and feature extraction were considered. The self-organizing map (SOM) was used to visualize the process state in [2]. An algorithmic variable selection approach is reported in [5]. Based on this, it is demonstrated how to find the on-line variables that best describe an off-line defined process phenomenon [3]. Another use for the variable selection technique is reported in [4], where the process control variables best explaining changes in process control success were searched for. The current work aims at creating a methodology for finding the variables that train a SOM in which the phenomenon defined by the user is presented as clear as possible.

References

[1] S. Laine, K. Pulkkinen, and S.-L. Jämsä-Jounela. On-line determination of the concentrator feed type at Outokumpu Hitura Mine. *Minerals engineering* pages 881–895, No. 8–9, August 2000.

[2] J. Rantanen, S. Laine, O. Antikainen, J.-P. Mannermaa, O. Simula, and J. Yliruusi. Visualization of pharmaceutical unit operations using self-organizing maps (SOM). J. Pharm. Biomed. Anal., pages 343–352, January, 2001.

[3] S. Laine. Combining Off-Line and On-Line Information in Process Study Using the Self- Organizing Map (SOM) Proceedings of IEEE Mountain Workshop of Soft Computing in Industrial Applications, ed. Embrechts J., VanLandingham H., Ovaska S., Blacksburg, USA, pages 71–76, June 2001.

[4] S. Laine. State Based Process Study Using the SOM and a Variable Selection Technique in Evolving Solution with Neural Networks. In Proceedings of International Conference on Engineering Applications of Neural Networks, ed. R. Baratti, J. Caete. Cagliari, Italy, pages 15–22, July 2001.

[5] S. Laine. Finding the Variables of Interest. *Minerals engineering* (accepted).

16.3 From fault detection to fault identification with normal condition process data

Fault detection and fault identification are the two problem types encountered when analyzing industrial process data. Solving the fault detection problem in the presence of normal condition data only — in the framework of novelty detection — one can train a process model of the normal state and detect any deviations from this model as being abnormal. On the contrary, if fault data is available, fault identification problem can be solved in the classification setting, where process data are the symptoms and the faults are the classes. One interesting problem class that often occurs in practice is the need to solve the fault identification problem with normal condition data only, and perhaps a few labeled samples of the faults or some general prior knowledge of the fault types. We are interested in solving this problem that in fact falls in between the standard treatments of fault detection and fault identification problems.

In [1], we compare trained models from normal process condition data by studying the usefulness of residuals in the problem of fault identification. More specifically, we study the generation and visualization of residuals for detecting and identifying novel, unseen faults using autoassociative models learned from process data. Least squares and kernel regression models are compared on the basis of their ability to describe the support of the data. Theoretical derivations are used to show that kernel regression models are more appropriate in approximating the normal condition model. Moreover, experiments on vibration and current data from an asynchronous motor confirm the theory and yield more meaningful results when testing the models with some known fault classes.

References

[1] I. Díaz and J. Hollmén. Residuals Generation and Visualization for Understanding Novel Process Conditions. Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN'02) (to appear).

16.4 Decision models for computerized decision support

Several computerized decision support system prototypes have been developed and summarized in [1], where decision making problem formulation of failure management in safety critical processes was studied. The main application area were the nuclear power plants. Decision support in both the control room and maintenance were covered. The methodological frame of the models consists of decision theory, knowledge engineering and object-oriented modeling. One of the models, the use of decision analysis methodology in maintenance problems, is presented in [2].

The problem in the control room is how to produce instructions for the personnel responsible for the plant in situations where faults have occurred at various stages. In maintenance mainly off-line support is needed. Methodology and models based on application knowledge are developed to meet the functional requirements of decision support systems. Prototypes are built to verify the usability of the methodologies and models. The tools are demonstrated and tested by solving concrete decision problems. One of the key issues is the methodology developed in strategy planning, including both generation and selection of strategies.

After these works mentioned the objective has been to build decision models for certain applications, mostly for practical purposes, and also try to find out more general decision making principles. This approach leads easily to single case studies that are difficult to generalize. The large amount of possible methodologies and the narrowness of application areas are also known difficulties.

To find out general principles from separate case studies, to formulate more comprehensive decision concepts, and to build more general decision models are difficult tasks. As results of these studies tested models and concepts are got, but the evaluation of them is difficult, because no competent measures exist for such purposes. The only really clear result is the decision support achieved in each particular case.

How to utilize data analysis in computerized decision support systems has been outlined. The methodologies used in Finnish industry, including chemical industry, nuclear and conventional power production, has been collected and analyzed in an interview study. A decision model for state monitoring has been developed. This model utilizes fuzzy methodologies, traditional calculation techniques and heuristics.

An old decision case that has been analyzed with rule-based methodologies in [1] has been solved with multi-criteria decision analysis method in [3]. Comparisons with the elder case has been made in the analysis. The problem is to choose the right control action in a situation where a leak has appeared in the primary circuit of a BWR nuclear power plant.

References

[1] M. Sirola. Computerized decision support systems in failure and maintenance management of safety critical processes. VTT Publications 397. Espoo, Finland, 1999.

[2] K. Laakso, M. Sirola, and J. Holmberg. Decision modelling for maintenance and safety. International Journal of Condition Monitoring and Diagnostic Engineering Management. Birmingham, England, July 1999.

[3] M. Sirola. Applying decision analysis method in process control problem in accident management situation. International Conference on Systems Science. Wroclaw, Poland, September 2001.

16.5 The other projects

Fault detection in continuous casting. Continuous casting of steel has been investigated in a project with Suomen Perusmetalli. The casting process is a large scale industrial method for producing steel slabs for further refinement, e.g., for hot rolling. During the casting, surface ruptures may appear on the steel slab. They degrade the product quality, and sometimes they may even cause a breakthrough of liquid steel that leads to a long maintenance break at the casting plant. The measurements monitoring the casting process are done using thermocouples inside the mold. These temperature measurements have been analyzed using SOM-based visualization tools in order to find efficient features for rupture and breakthrough warnings [1]. At the moment, satisfactory features have been discovered and used to build a classifier to detect ruptures. The state of the process can be indicated using the trajectory of measurements on the visual SOM display. The SOM can, thus, be used for on-line monitoring of the casting process.

Analysis of hot rolled steel production. During the period 1.2.1997–31.1.2000 "Application of Neural Network based Models for Optimisation of the Rolling Process", NEUROLL, was carried out. The project was financed by the European Union under the program Brite-Euram III. The rationale of this project was to apply intelligent computation methods, especially neural networks to the steel strip rolling, the ultimate goal being to improve quality and throughput of the products. From the product point of view, the project was concentrating on such dimensional properties as width, thickness and flatness; on the material properties, especially phase transition; and on the surface quality control.

The research carried out in the NEUROLL project was focused on two subprojects: (1) The detection of complex relationships between measured or calculated process parameters and input/output variables and (2) analysis of the surface quality of the hot rolled steel strips. The SOM based data analysis was carried out for visualization and classification of various measurements and data dependencies. The study has concentrated on finding correlations between quality parameters (e.g., width and thickness deviation and surface defects) and other process parameters using traditional correlation analysis and methods based on the SOM. [3, 2] As an example, Figure 16.2 below shows the dependence of the quality parameters on the chemical content, including the temperature measured after the last stand, could be seen that the deviation of the temperatures has a clear influence on the thickness and width deviations. [4]

References

[1] J. Himberg, J. Ahola, E. Alhoniemi, J. Vesanto, and O. Simula. Pattern recognition in soft computing paradigm. Volume 2 of FLSI soft computing series, ed. Nikhil R. Pal, chapter The Self-Organizing Map as a Tool in Knowledge Engineering. World Scientific, 2001.

[2] J. Parviainen. Data Mining for Finding Surface Defects in Steel Strips. Master's thesis, Helsinki University of Technology, 2000.

[3] J. Ahola. Analysis of Mechanical Properties of Hot Rolled Steel Strips Using SOM-based Data Mining. Master's thesis, Helsinki University of Technology, 2001.

[4] Application of Neural Networks based Models for Optimisation of the Rolling Process (NEUROLL). Final Technical Report, Contract No. BRPR-CT96-0342, Project No. BE96-3061.

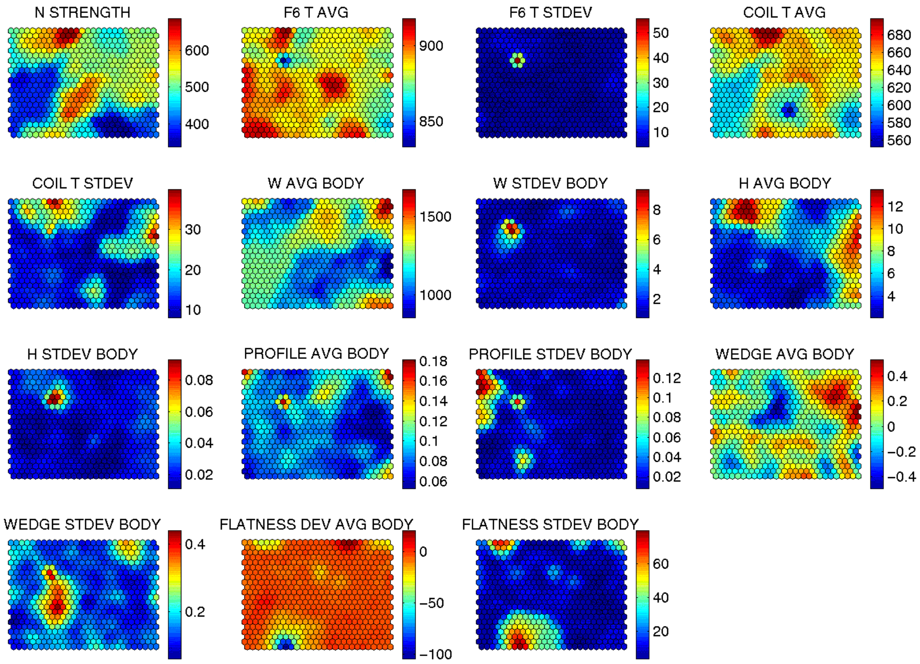

Figure 16.2: Temperature vs. quality parameters (micro-alloyed steels).