AGH

University of Science and Technology in Krakow

Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering

Department of Automatics and Robotics

Doctoral Dissertation

Tian Cong

Statistical reasoning analysis of fault occurrences in industrial applications

Discipline:
Automatics and Robotics
Supervisor:
Jerzy Baranowski Ph.D
Assistant Supervisor:
James R. Ottewill Ph.D

Krakow 2020

Akademia Górniczo-Hutnicza

im. Stanisława Staszica w Krakowie

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

Katedra Automatyki I Robotyki

ROZPRAWA DOKTORSKA

Tian Cong

WYKORZYSTANIE WNIOSKOWANIA STATYSTYCZNEGO DO ANALIZY WYSTĘPOWANIA USTEREK W ZASTOSOWANIACH PRZEMYSŁOWYCH

DYSCYPLINA NAUKOWA:

Automatyka i Robotyka

PROMOTOR:

Dr. hab. in˙z. Jerzy Baranowski

PROMOTOR POMOCNICZY:

Dr. James R. Ottewill

Kraków 2020

1 Goal of the thesis

As modern industrial plants are instrumented with a large number of sensors, advanced monitoring algorithms are required to extract actionable insights from the vast quantities of process measurements. The goal of this thesis is to develop monitoring algorithms for detecting abnormal operation from industrial processes, additionally providing a workflow which is able to handle the monitoring of processes with different range of complexity. In addition, the developed algorithms should be practically relevant, considering the ease of data management, computational efficiency, the scalability and complexity of algorithms and other implementation issues. The monitoring statistics should be traceable, interpretable and visualisable for end users.

2 Motivations

The thesis consists of 7 chapters. Chapter 1, 2, 6 and 7 are Introduction, review of Process Condition Monitoring, Experiment Case Studies and Conclusion, respectively. Chapter $3 - 5$ are technical chapters. As this thesis aims to solve the monitoring issues with various complexity, each technical chapter is motivated by a certain application scenario.

- Chapter 3 focuses on the adaptability for monitoring a single-mode process. The adaptability is required when there are only a small amount of measurements, particularly if these measurements are insufficiently representative. In such cases, it is desirable that the designed algorithm can adjust its monitoring thresholds by continuously incorporating the information from the incoming process data. In addition, simultaneously inspecting multiple process variables might be inefficient and cumbersome for engineers and operators. Thus, a single monitoring indicator extracting insights to multiple process variables should be designed to assist operators in learning about the health state of processes. Moreover, the derived monitoring thresholds should not be restricted to Gaussian distributions, thus the capability of handling the non-Gaussian cases is taken into account.

- Chapter 4 focuses on the ease of data management in multi-mode processes. Particularly in the cases where the number of operation modes is not known in advance, automating the data clustering is demanding. Due to the varying loading conditions or production regimes, the recorded data would be a mix of various operating modes. There is a need to separate them according to the operating modes such that the characteristics of each individual mode can be further analysed.

  However, the clustering results might be sensitive to the initialisation of the hyperparameters of clustering algorithms, for instance, the Dirichlet Process (DP). To make the clustering results reliable and accurate, it is worth to investigate how to properly set the initial values of hyperparameters. Moreover, the operating modes in the data base might be enriched over time. The monitoring model trained with previous historical data might be insufficient. In an on-line monitoring manner, it is desirable to identify the occurrence of new healthy modes, then re-train the monitoring model by incorporating the data from new modes.

- Chapter 5 focuses on the fault detection from a multimode processes with a model-based monitoring model. In a process, various operating modes might have the same or similar steady-states, but with distinct dynamics. Monitoring algorithms that do not take dynamics into account would fail to distinguish such modes. The Field Kalman Filter (FKF) is a model-based Bayesian algorithm, being capable of simultaneously estimating the state, system parameter and noise parameter. This advantage can be applied to differentiating various operating modes both deterministically and stochastically.

  One of the barriers of applying this model-based method is the process modelling of modern industrial processes. The modelling difficulty arises from the growing complexity of industrial plants. For example, it is challenging to describe the physics of the intricate interlinked equipment with first-principles. With data analytics, the process models can be obtained using historical data. In addition, the local dynamics existing in the healthy data may be well described by linear discrete models.

  The FKF in the application of distinguishing various known operations is based on Bayesian statistical decisions which are interpretable and visualisable. However, Bayesian methods for anomaly detection are limited to the known process operation. Therefore, unknown operation, such as the occurrence of faults or new operating modes, might be undetectable. To extend the FKF for anomaly detection, the ability of inferring the anomalies should also be included in the algorithm design

2.1 Contributions:

The main achievements are shown as the flowchart in Fig. 1. It is possible to choose a suitable monitoring approach according to the number of operating modes in the monitored systems. Each round box (e.g. modules and algorithms) in this flowchart corresponds to a contribution of this thesis. End users can incorporate these contributions as a whole or as individual sections into their own monitoring framework.

2.1.1 Detailed contributions in Chapter 3

- Develop the BaFFle algorithm with adaptive fault detection, The BaFFle is a heuristic and low-complexity algorithm, which can adjust its monitoring model by continuously involving the information from the preceding data.

- Develop a method for applying univariate control charts in monitoring multivariate processes. The Principal Component Analysis (PCA) technique is used for extracting uncorrelated

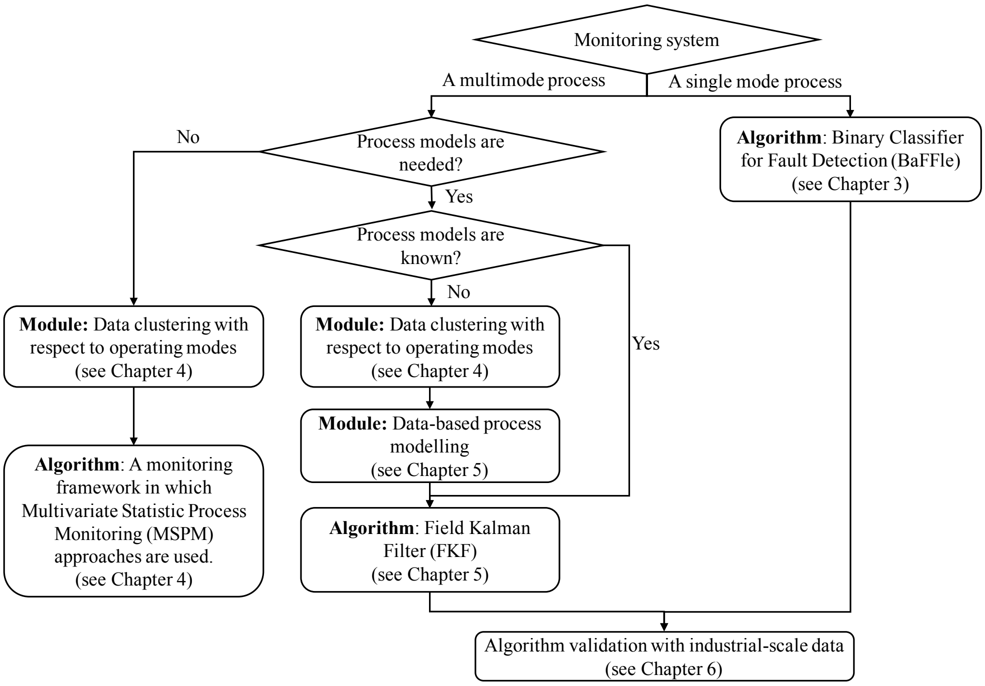

Fig. 1: Main achievements in this thesis

features from multivariate data so as to have multiple unbiased univariate control charts. To fuse the decision across individual control charts, a majority voting strategy is adopted;

- Employ the nonparametric method, Kernel Density Estimation (KDE), to cater to non-Gaussian distributions;
- Create a mechanism of warning and detection, where the warning indicator is used for adjusting control limits while the detection indicator is used for determining whether a process is healthy or not;

2.1.2 Detailed contributions in Chapter 4

- Apply the DP-GMMs for data clustering, in which the Markov Chain Monte Carlo method is utilised to infer the number of patterns existing in the healthy process data;
- Investigate the impact of the initial hyperparameters of the DP-GMMs on the clustering results;
- Propose a method to minimise the impact from improper initialisation;
- Validate the proposed method using simulation examples;
- Propose an on-line monitoring framework. In this framework, additional to classifying the unlabelled historical data with respect to various operating modes, the DP-GMMs are also used in the situation that on-line data are detected as new operating modes. In this case the

historical data and on-line data are combined and clustered using the DP-GMMs approach in order to update the information of normal operation required in training monitoring models.

2.1.3 Detailed contributions in Chapter 5

- Demonstrate the ability of the FKF in fault classification and isolation with a simulation model. Investigate the data-based process modelling in the absence of prior knowledge of process industries. As the FKF monitoring model for a multimode process is a set of state-space models, the Multivariate Autoregression State-Space (MARSS) approach is employed in this thesis;

- Extend the original FKF algorithm for anomaly detection by designing a novel unified monitoring indicator;

- Develop a workflow for systematically using the FKF in the industrial applications of anomaly detection and mode identification.

2.1.4 Detailed contributions in Chapter 6

- Validate the BaFFle algorithm for fault detection using industrial-scale process data. The fault detection results have presented the effectiveness and interpretability of the BaFFle algorithm. The effectiveness of the dynamic control limits in reducing false and missed alarms have been demonstrated in the comparison with constant control limits;

- Apply the proposed FKF workflow for monitoring a multimode process. The effectiveness of proposed workflow has been proved using the PRONTO benchmark data. This dataset consists of hybrid data from a range of sensors distributed across a pilot-scale multiphase flow facility. The experiment results have shown that the use of Bayesian statistic decisions can differentiate various operating modes, and that anomalies can be detected with the unified monitoring indicator. In addition, the comparison experiment has highlighted that the FKF monitoring model can achieve a relatively low false alarm rate, and outperforms the two selected approaches at mode identification.

2.2 Summary

This thesis focused on systematically designing novel monitoring methods and algorithms for detecting abnormal behaviours from measurement data. This includes considering processes with a variety of complexity such that, with the newly designed approaches, monitoring systems would respond to the occurrence of faults and anomalies with more reliability and efficiency. The experiment results show that these the proposed monitoring methods can improve the detection performance, particularly shortening the detection time and reducing the false and missed alarm rates.

References

[1] J. Baranowski, P. Bania, I. Prasad, and T. Cong, "Bayesian fault detection and isolation using field kalman filter," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, pp. 1–11, 2017.

[2] T. Cong and J. Baranowski, "Binary classifier for fault detection based on gaussian model and pca," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 1317–1323, 2018.

[3] ——, "Binary classifier for fault detection based on kde and pca," in *2018 23rd International Conference on Methods & Models in Automation & Robotics (MMAR)*. IEEE, 2018, pp. 821–825.

[4] R. Tan, T. Cong, N. F. Thornhill, J. R. Ottewill, and J. Baranowski, "Statistical monitoring of processes with multiple operating modes," *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 635–642, 2019.

[5] R. Tan, T. Cong, J. R. Ottewill, J. Baranowski, and N. F. Thornhill, "An on-line framework for monitoring nonlinear processes with multiple operating modes," *Journal of Process Control*, vol. 89, pp. 119–130, 2020.

[6] T. Cong, R. Tan, J. R. Ottewill, N. F. Thornhill, and J. Baranowski, "Anomaly detection and mode identification in multimode processes using the field kalman filter," *IEEE Transactions on Control Systems Technology*, 2020.