Data-Driven algorithms for fault detection and diagnosis in industrial process

M. EL KOUJOK, and M. AMAZOUZ
Industrial Systems Optimization Group, CanmetENERGY, Varennes, QC, Canada

Abstract— *Data-driven methods have been recognized as useful tools to extract knowledge from massive amounts of data. However, their use for process operation monitoring and fault diagnosis is still confronted by some challenges. This work aims to facilitate the development and the use of models for fault detection and diagnosis (FDD). To do so, a Matlab-based software has been developed. Latent variables methods such as principal component analysis (PCA) and projection to latent structures (PLS) are integrated into the software, along with appropriate control charts and interpretable plots of variables causing faults. The software enables loading of historical data from various sources, automatic building of models that describe the normal operation, and prediction of quality variables. The software can easily be connected online for continuous FDD of a process. Such tool serves as a decision support for process operators. The usefulness and the accuracy of the tool for FDD is demonstrated with the Tennessee Eastman testbed.*

Keywords: Data driven methods, FDD, chemical process.

1. Introduction

Quantitative data-driven methods have recently received considerable attention from chemical industries, due to the huge amounts of data and the effectiveness of its analysis and interpretation. A team of experts is required to put in place these types of methods and assist the industry in analyzing historical and real-time data. It is however cost intensive for a manufacturer to derive value from data. This work is the first step in developing a Matlab-based FDD software tool that can assist engineers to continuously monitor and rate the performance of a process and perform online FDD. The tool is a dashboard that illustrates fault detection time and the isolation of abnormal events in real time. By using this tool, the operating costs for a plant will be decreased because it will no longer be necessary to hire external consultants.

A wide range of data-driven algorithms can be found in the literature to support the design of an advanced FDD tool. Through the use of machine-learning based model, these algorithms transform data into knowledge. A process can then monitored using the constructed data driven model. Questions such as, how to deploy these methods to help the engineer must be answered. Today, one of the most challenging tasks that chemical engineers face is effective monitoring of entire chemical process complexes, with many interconnected units of operation (i.e. maintaining normal and optimal process operation, as well as ensuring component balances, product quality production rates, and environmental regulations compliance, etc.). This paper presents a developed tool based on appropriate data-driven methods for FDD. A detailed study guided the choice and implementation of two appropriate data-driven methods: (1) Principal Component Analysis (PCA), which focuses on the study of one data block for which all process variables are monitored; and (2) Projection to Latent Structures (PLS), which can serve as a powerful tool for monitoring key performance indicators (KPI). An advantage of these methods (PCA and PLS) is the projection of the original variables onto a latent subspace: latent variables will be monitored in a reduced dimensional space, thus preventing the user from having to select variables. All of the variables for a complex chemical plant can then be monitored. In addition, the use of alternative supervised classification fault diagnosis methods such as support vector machine (SVM) or Neural Network (NN), are not always adequate because it is difficult to identify the mode of abnormal operations ahead of time in real applications [1]. However, once PCA and PLS models are built on good historical data reflecting normal process operation, they can then be used to monitor and diagnose new faults, using indices and reconstruction-based multivariate contribution analysis, respectively [2].

2. Using the FDD Tool

The interface of the developed Matlab-based FDD tool is easy to navigate (Figure 1). Engineers/operators first load historic data from normal process operation. The PCA does not impose any restrictions with respect to the variables that can be employed to detect and isolate a new fault. The user clicks on a single button to build the PCA. For the PLS, a list containing the names of variables can be added to the platform, making it possible to select the input variables and output variable (KPI). The user can also construct the PLS model by a single button click. Several monitoring indices and control limits are integrated into the platform tool. In addition, the tool automatically considers the PLS output (KPI) as one of the monitoring indices and the user can establish its limits. A process is considered out-of-control or faulty if one of these monitoring indices falls outside the established control limit. Following fault detection, the

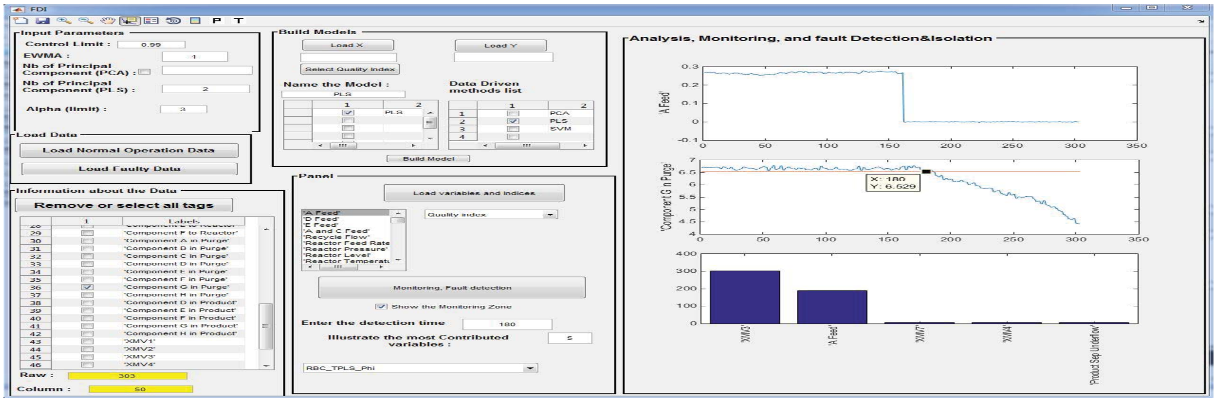

Fig. 1: Dashboard of the developed tool.

tool will display the variables that contribute the most to the detected fault and this is important in helping the engineer/operator find the sources of the detected anomaly. Measures can be taken to recover from the process fault and ensure that the operation remains normal or optimal. The tool that was developed has been applied to a sophisticated chemical process case study – the Tennessee Eastman (TE) benchmark problem [3].

3. Case study and results

The TE is a plant-wide industrial process, proposed as a benchmark for the Eastman Chemical Company. The plant consists of five main units: a two-phase reactor, a condenser, a recycle compressor, a liquid-vapor separator, and a product stripper. Two products are obtained from four reactants (A, B, C and D) in the process. The TE process consists of 11 controlled valves, and 41 measured variables that include temperature, level, pressure flow rate, and concentration. A TE simulator generated industrial data to help evaluate the two studied FDD techniques. The simulator can also generate different types of faults. To show the applicability of the proposed tool, normal operation data were generated to build the PCA and the PLS models and two faults (Table 1) were introduced separately to test each model.

Table 1: Table 1

Fault	Description	Type
Mode 1	Reactor cooling water inlet temperature	Step
Mode 2	Reactant ‘A’: feed loss	Step

FDD with PCA for fault mode 1: The user loads the normal operation data and the labels of each measured variable. The PCA can then be built to monitor the TE process. Process FDD was started with a click. Once fault mode 1 was seeded, the fault was detected within six minutes by one of the monitoring indices, and a bar chart presented the variables that contribute most to the fault. The result showed that the valve controlling the reactor cooling flow and the reactor temperature are most influenced by the fault. Based on this result, we could conclude that the reactor cooling water system was most likely faulty.

FDD with PLS for fault mode 2: Fault mode 2 had an impact on the production rate (a KPI), which started to decrease from its targeted value (Table 1). To detect this, a user must build a PLS by selecting the production rate as the output (KPI), and the remaining variables as input. The KPI deviation from its target can be easily detected by establishing a limit. The tool then uses the parameters of the PLS model to create a bar chart illustrating the variables that are most affected by this fault. The longest bar represents the position of the control valve for reactant ‘A’ and the second longest the flow rate of the same reactant. As a result, the user can identify the source of the fault to be the feed loss for reactant ‘A’.

4. Conclusion and future work

This work presents a Matlab-based software tool in the early stages of development, which can be used to detect and diagnose an abnormal chemical process operation. The benchmark case study of the TE process was used to highlight the usefulness of the tool. It can help engineers and operators to detect and isolate faults when there is no prior knowledge of faults. The methods integrated in this tool are appropriate for detecting and isolating a single fault at a time. Future work will focus on developing the tool to allow for the detection and isolation of multiple faults.

ACKNOWLEDGMENTS This work was supported by Natural Resources Canadas’s PERD program (NRCan).

References

[1] J. MacGregor, A. Cinar, “Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods,” *Computers & Chemical Engineering.*, vol. 47, pp. 111–120, 2012.

[2] G. Li, S.J. Qin, J. Yin-Dong and D. Zhou,“Total PLS based contribution plots for fault diagnosis,” *Acta Automatica Sinica.*, vol. 35, pp. 759–765, 2009.

[3] J.J. Downs, E.F. Vogel, “A plant-wide industrial process control problem,” *Industrial Informatics, IEEE Transactions on.*, vol. 9, pp. 2226–2238, 2013.