Demonstrator of high-tech system-diagnostics with operational data

Yonghui Li* and Emile van Gerwen

ESI, Eindhoven, The Netherlands

*yonghui.li@tno.nl

Challenges and opportunities

In big-data era, high-tech companies are encountering opportunities of using the emerging data to improve their system life cycle and to define new business cases by transforming themselves into service providing companies. Moreover, the massive amount of data generated by the high-tech systems also bring the opportunities to deal with the previously unseen challenges, such as the significant increase in system complexity, customization, continuous evolution and diversity of operational environment.

Main challenges emerging in *"how to benefit from data"* concern: 1) the automated extraction of relevant data insights (e.g., features and patterns) from a large and diverse amount of data streams, and 2) the interpretation of the insights and their integration into the existing system engineering.

ESI's contribution

ESI has both the knowledge and experience of applying data science and system engineering in high-tech system domain. We are developing a demonstrator to explore and address the challenges of bridging these two parties (i.e., data science and high-tech system engineering). Our demonstrator sketches out the landscape and integration principles of knowledge-assisted data analysis techniques applied to a variety of data streams available within high-tech industry.

Demonstrator on system diagnostics

As an industrial showcase, our demonstrator presents semi-automatic identification of the main root cause of a factory (system-of-systems) throughput degradation issue through a guided and deep-dive analysis to the level of a machine-specific component (a software task). Figure 1 shows the analysis flow of our demonstrator, where multiple data science techniques are applied for different data streams for identifying the possible root-causes. The detailed explanation of the technologies in the demonstrator is given in next section.

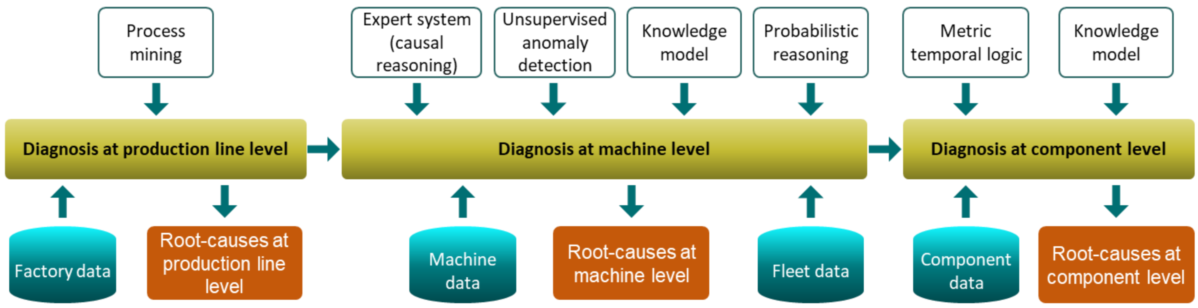

Figure 1: The analysis flow of our demonstrator on system diagnostics.

Deployed technologies in the system diagnostics demonstrator

The overall intelligence and added value of the developed demonstrator lies in the integration of multiple state-of-the-art data science techniques for analysing data streams at various diagnostics levels. Figure 2 presents a technological overview of the demonstrator.

Data analysis Technique / Data Source	Process mining	Anomaly detection / unsupervised learning	Expert system / causal reasoning	Probabilistic reasoning	Metric temporal logic	…
Production line data (system of systems)	1) Identifying process flow bottlenecks					
Machine data		2) Unsupervised identification of anomalies	3) Expert knowledge guided analysis to identify likely causes			
Fleet data				4) Exploiting statistics to rank likely causes		
Machine component data					5) Identifying blocking dependencies in interacting processes	
…						

Figure 2: Technological overview of the demonstrator, where all data analysis techniques can be applied to all data sources (light green boxes). However, in the demonstrator, we have selected across combinations (dark green boxes) for specific diagnostic challenges.

The demonstrator is currently built around 5 diagnostic challenges, each of which uses different types of data streams and analysis techniques. We note that the demonstrator is work-in-progress and more analysis techniques will be added in due time.

1) **System-of-systems data from a production line & process mining**.
   A production line usually consists of multiple machines operating in a system-of-systems manner. These machines cooperatively process the same objects, and generate operational data. Process mining [1] is employed to analyse the operational data and derive the process flows of objects on the production line. The performance bottlenecks (e.g., object distributions across machines and/or specific machines) can be identified by analysing the extracted flows (see the screenshot of the demonstrator in Figure 3). Given the identified performance bottlenecks, next-step analysis, e.g., for specific machines, are recommended.

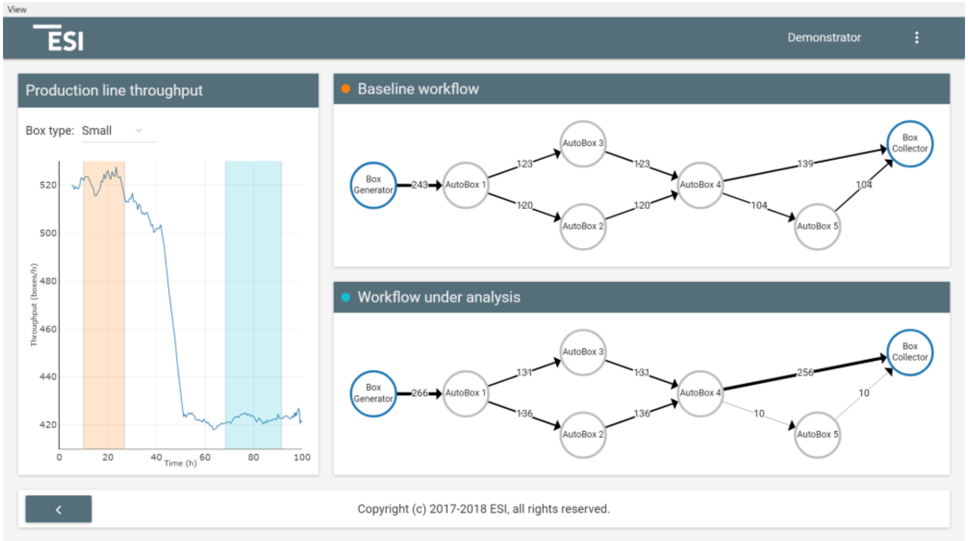

Figure 3: Production line performance bottleneck analysis using process mining.

2) **Machine data & anomaly detection**. To diagnose a specific machine issue and narrow down the space of causes, we use the operational data generated by the machine components. Basically, anomalies are identified based on the operational data. This is achieved using unsupervised learning techniques, which are able to detect anomalies without efforts from domain experts. These anomalies shown on the screenshot of the demonstrator in Figure 4 are provided by Yazzoom.

Figure 4: Unsupervised anomaly detection based on machine data.

3) **Machine data & expert system.** The detected anomalies are further linked to the system engineering knowledge, captured as causal relationships about the components [2] (see the screenshot of the demonstrator in Figure 5). The likely root-causes of the machine issue are hence obtained according to the causal relationships and the anomalies. This method represents a practical application of knowledge-assistant data analysis in industry.

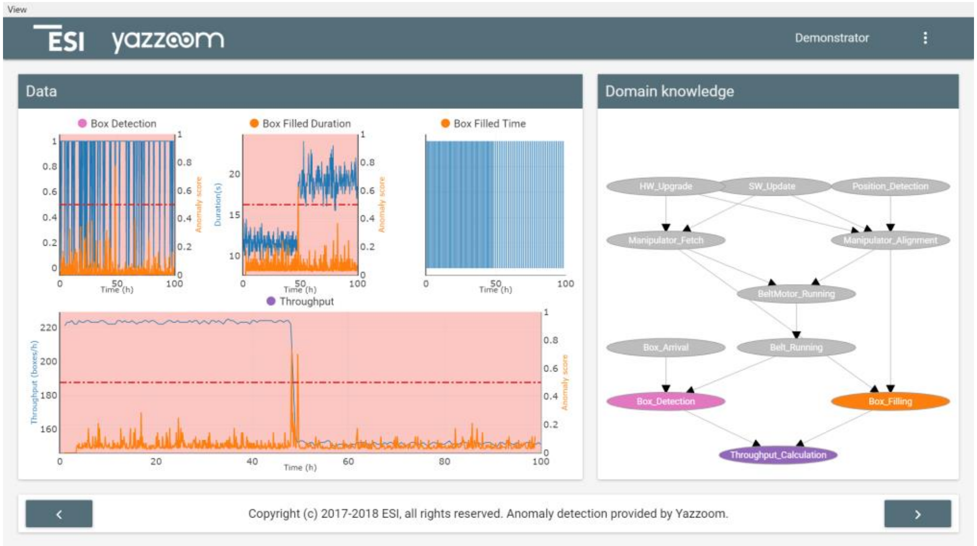

Figure 5: Root-cause analysis guided by the system engineering knowledge.

4) **Fleet data & probabilistic reasoning.** The operational data of a fleet of machines further provide evidence whether a likely cause identified on one machine also leads to the same issue on other machines. To reason about such fleet data and knowledge, probabilistic reasoning [3] is used to statistically rank the likely causes and recommend further deep-dive diagnosis of the machine component (see the screenshot of the demonstrator in Figure 6).

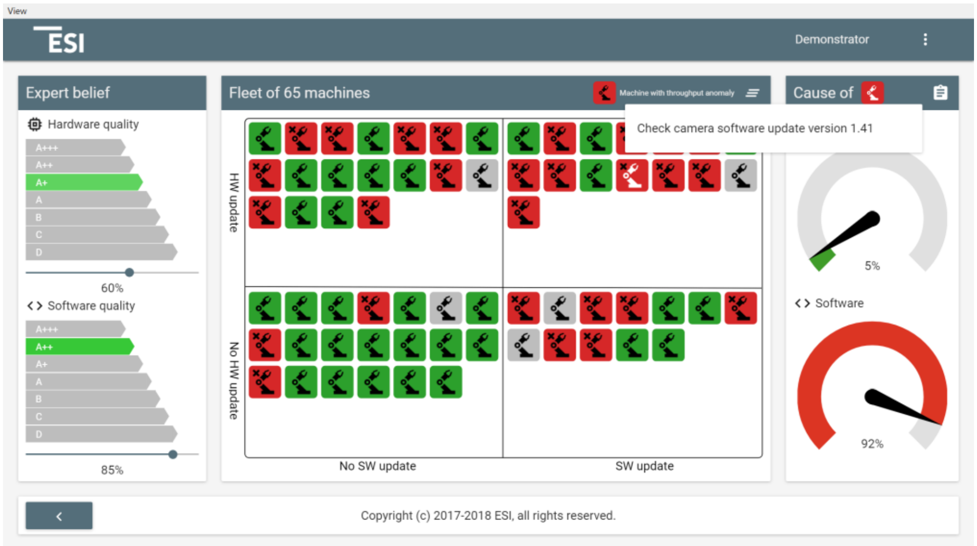

Figure 6: Probabilistic reasoning based on fleet data.

5) **Machine component data & metric temporal logic.** The diagnosis of a machine component relies on its operational data, such as the log events. The failures or errors of the component are identified by verifying its performance metrics, such as latency or throughput requirements. This is achieved using the metric temporal logic (MTL) [4], which formally verifies the performance requirements with the log data of the component. If the requirements are not satisfied, failures or errors are visualized, with an appropriate recommendation for actions. Note that the specification of the requirements is described in a domain-specific language (DSL) that engineers can easily understand. The formal language (i.e., the metric temporal logic) backing the analysis is invisible to the engineers.

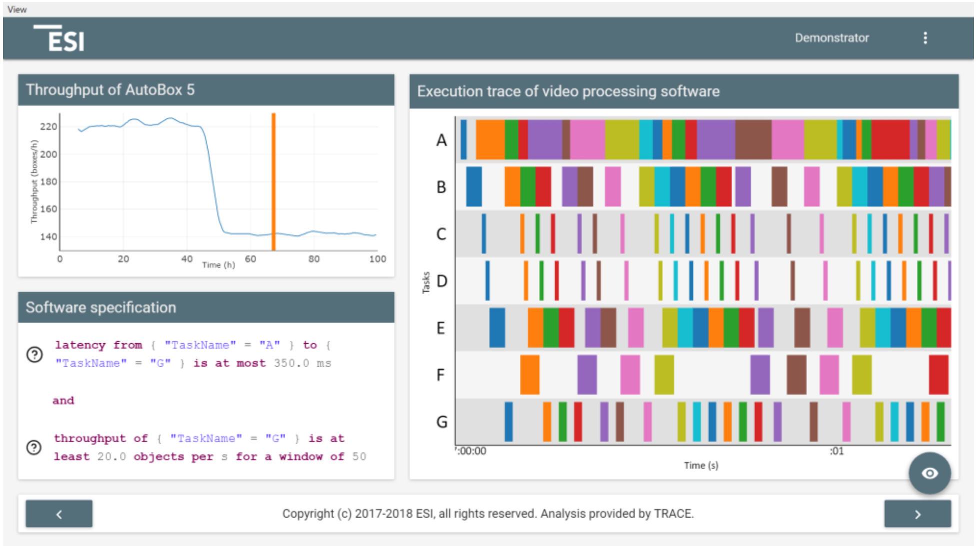

Figure 7: Failure/error identification using metric temporal logic analysis of machine component data.

CONTACT

For a LIVE DEMO or further information, please feel free to contact Yonghui Li (yonghui.li@tno.nl) and Emile van Gerwen (emile.vangerwen@tno.nl).

ACKNOWLEDGEMENT

The work was carried out as part of the ITEA3 REFLEXION project under the responsibility of ESI. The REFLEXION research is supported by the Netherlands Organisation for Applied Scientific Research TNO and Netherlands Ministry of Economic Affairs. We thank Tomas van Oyen from Yazzoom for providing the unsupervised anomaly detection of the demonstrator.

REFERENCE

[1] W.M.P. van der Aalst, “Process Discovery: An Introduction”. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, pp 125-156, 2011.

[2] J.J. Rooney, L.N. van den Heuvel, “Root cause analysis for beginners”. Qual. Prog. 37(7), pp 45–53, 2004.

[3] D. Heckerman, J. B., “Causal independence for probability assessment and inference using Bayesian networks”. *IEEE Transactions on SMC – A*, 26(6), pp 826–831, 1996.

[4] M. Hendriks, J. Verriet, T. Basten, B. Theelen, M. Brassé, L. Somers, “Analyzing execution traces - critical path analysis and distance analysis”. *Int. J. on Software Tools for Technology Transfer,* 19(4), pp 487–510, 2017.