Research Plan: Simulating Flexible Assembly System Process Traces for the Purposes of Process Mining and Machine Learning – DRAFT

Tero Keski-Valkama

May 5, 2016

Contents

1 Introduction

Flexible Assembly System is a modular and reconfigurable assembly and tooling workshop with a focus on small and medium-sized batches of varying products.

Using a Flexible Assembly System to assemble batches of products generates an electronic event trace which is logged. The events in the event trace are typically at least partially agnostic to the assembly process being performed, and therefore the events do not include tokens connecting the event to a specific final product instance. The process traces formed by logs are therefore uncorrelated.

Simulating these kinds of event traces provides valuable material for learning systems which can be used to derive implicit, behavioristic assembly process models and to detect deviances. Deviances which are not detected with current methods with alerts and fault codes are useful for detecting unexpected failure conditions, such as human errors and cyberattacks.

It is in principle possible for an automatic system to detect features in the given material in an unsupervised fashion at least when a human is capable of doing so. We will use multimodal visualization to represent the generated process traces to discover features interesting as machine learning targets.

Modern industrial plants are gaining advanced data acquisition capabilities rapidly because of Internet of Things technologies and open, standard APIs. Malfunctions are a significant cause for expensive downtime in flexible assembly systems. Traditionally malfunctions are detected through specific and pre-designed fault codes and diagnostic metrics. However, especially in the context of predictive maintenance the fault codes are lacking as they typically signal only at the point where downtime actually happens.

Preventive maintenance is done by periodical scheduling and at the discretion of human operators when something seems to start breaking apart. Often the human operators periodically check the process metrics and event logs to see any discrepancies and anomalies. There is a lot of potential in automatic anomaly detection, as the anomalies can be observed instantly as they happen and not only after the human operator checks the logs and metrics the next time. Automatic anomaly detection can also potentially detect anomalies that would otherwise not be immediately evident for human operators.

Modern flexible manufacturing and assembly sites are very heterogeneous and different end products require different workflows and sequences in the system. Operation that is normal for one site might be considered anomalous for another. However, there are some general characteristics that are common for all sites with approximately similar production processes, especially for

distinct components of the system.

I got first hand experience in the problem field and potential solutions in #IndustryHack #HackTheFactory competition in Fastems in May 22th – May 24th, 2015[1]. My team won the competition with a concept utilizing a trivial and naive Markov chain model of the process with crude anomaly detection capability. There is a clear need of such solutions, and considerable improvement of the concept is possible through researching novel application of modern machine learning in the context of industrial automation system anomaly detection. The Markov chain model was roughly equivalent to a previously studied DFA model[2] which was ultimately deemed inadequate for situations where log messages of parallel processes are interleaved together into the same process trace. DFA and Markov chain models assume there is only one process instance reflected in the logs, and cannot learn to separate different processes. Additionally it might be impossible to manually multiplex log messages to separate state machines as there might be log messages that are shared for several separate processes.

These kinds of interleaved, symbolic event logs, or process traces, are common for many fields. For this research, a Flexible Assembly System is taken as a case for modelling for the purpose of enabling research and development of learning systems for anomaly detection and predictive maintenance.

1.1   Core Questions

The core questions about the existing literature are:

1. How are assembly systems described and modelled?

2. What methods are there to model and simulate assembly processes and related faults?

3. What methods are there to visualize and represent event logs with or without timestamps?

4. What are the relevant keywords and terms to describe this problem space?

2   Problem Statement

Creating new types of process mining and learning systems for anomaly detection in Flexible Assembly Systems requires data in the form of process traces. There is not enough publicly available data collected from real Flexible Assembly Systems, and it is challenging to gather a representative corpus

of real process logs with anomalies. The nature of anomalious traces is such that a certain kind of anomaly rarely reoccurs. Each anomalous event is unique and unexpected. This means that in real process traces a specific kind of anomaly correlates strongly with the preceding trace, as it is an unique event that has never occurred in other contexts. This make such data not representative in respect to anomalies. It is not feasible to train and validate learning systems against a dataset with only a few non-recurring anomalies.

A simulated model of a Flexible Assembly System is required to generate appropriate event streams and to simulate failure modes. These kinds of simulations do not exist at the moment in academic literature. While process deviances in assembly systems are described in the literature, there are no published simulations of such. Even when simulations are used as a basis of creating industrial anomaly detection systems, such simulations and real plant data are not published[3]. This makes it difficult to compare methods against common benchmarks and reproduce research results.

The generated process traces are interleaved and the events are uncorrelated. Finding good methods for process mining such data is an open research problem.

3 Concept and Methodology

This research aims to create a process level discrete event simulation of a Flexible Assembly System capable of representing correct operation and different failure modes. The research consists of a literature review of Flexible Assembly System modelling and simulation, and of relevant failure modes.

A discrete event simulation is created to model the behavior of a representative Flexible Assembly System. The simulator supports defining system under simulation using a Process algebra Petri Net (PPN) notation. The symbolic event streams are generated and visualized in different ways to evaluate the potential for applying different methods to such data. Special attention is given to the possibility of extracting interleaved process traces out of such data. If a human is capable of seeing or hearing the interleaved patterns in the data, then it should in principle be possible for automatic systems also.

The generated process traces are visualized to show their general form to direct selection of data representation for learning systems and statistical analysis.

Different kinds of batches and operating conditions are used, and resultant process traces are visualized to show general characteristics of such

production runs and failures.

The initial model of the system under simulation is loosely based on a Youtube video[4] of a real Flexible Assembly System in Chrysler transmission assembly line. The work steps are catalogued and timed, and a discrete event simulation is developed to roughly reflect that kind of production setting. The system is described and configured into the simulator using a Process algebra Petri Net. Faults from existing literature are added to the simulation to produce process deviances.

The process traces reflecting these process deviances are represented in visual and in audial form to assert that they have features that can in principle be detected by automatic systems.

4   Model Verification

The goal of the first phase of the research project is to create a source of interesting data for modelling and analysis. To be interesting, the simulation implemented should roughly correspond to typical conditions, although it does not need to exactly replicate any specific existing system. The simulation should be easily configurable to represent an arbitrary existing system of arbitrary complexity.

The represented process traces should somehow exhibit the failure conditions which would not be evident from explicit error codes in the logs. The process traces should be interleaved and uncorrelated, as they are so in the real systems.

The visualized process traces are inspected for features possibly detectable by a human, to identify potential targets for automatic methods.

5   Schedule and Status

The first part of the research is about simulating flexible assembly systems as outlined in this document. An article regarding the first part of the research will be ready for publication review and is en route for publication within 2016.

The second part of the research is about evaluating LSTM neural networks in anomaly detection regarding the material from the first part of the research. The work on the second part of the research starts with a new literature review and updated research plan. The second part of the research results in at least one published article.

The third part of the research is about methods of integrating anomaly

detection system in an issue management scheme, and results in at least one published article.

The three parts of the research are initially scheduled to take place within 2016-2018.[5]

The first article draft describes the fault-free operation of the Flexible Assembly System and its simulation and related existing literature. FAS Simulator initial version has been published in GitHub, simulating the fault-free operation of the described Flexible Assembly System. Literature review has been started for the systemic process faults of the Flexible Assembly Systems.

References

[1] "Cybercom team won the #IndustryHack #HackTheFactory Fastems - Summary," https://www.linkedin.com/pulse/cybercom-team-won-industryhack-hackthefactory-fastems-keski-valkama.

[2] F. Langer, K. Bertulies, and F. Hoffmann, "Self learning anomaly detection for embedded safety critical systems," *Schriftenreihe des Instituts für Angewandte Informatik, Automatisierungstechnik am Karlsruher Institut für Technologie. KIT Scientific Publishing*, pp. 31–45, 2011.

[3] N. Matarese, V. Colla, M. Vannucci, and L. M. Reyneri, "A procedure for the detection of anomalous input-output patterns," *Intelligent Data Analysis*, vol. 17, no. 5, pp. 737–751, 2013.

[4] Cars, "Chrysler transmission assembly line," 2014. [Online]. Available: https://www.youtube.com/watch?v=447sSED91v0

[5] "Kuha - uusi paradigma," https://goo.gl/ca48EG.