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

Tero Keski-Valkama

April 10, 2016

Keywords. Flexible Assembly System, Discrete Event Simulation, process mining, uncorrelated event streams, multi-modal representation

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. Flexible Assembly System was described formally by Donath and Graves [1] as a system consisting of a set of products each with a specified volume assembled on a workshop consisting of a fixed number of cells.

The assembly steps are performed in cells in parallel. Assemblies and components are transported to the cell where they are combined, tooled or inspected. Products are transported out of the cell for the next assembly step in another cell, or out of the system as end products. The work steps executed in the cells can be manual or automatic. There can be a central storage such as a shelf for

storing components, the intermediate assemblies between the steps and the end products waiting to be transported out of the system.

Flexible Assembly System work loads consist of small and medium sized batches where there can be some variation in the products based on customization and personalization. The Flexible Assembly Systems are modular and often composed out of independent modules from different suppliers. The Flexible Assembly System event trace consists of events received from all the separate modules of the system, and additional sensors and triggers added to the system in the integration phase or later.

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. A multimodal approach is convenient in data representation when we need to explore human pattern recognition capabilities. In specific, data visualization as audio allows a human to understand the data melodies, rhythms and possible canon-like features. This is why we will use multimodel visualization to represent the generated process traces.

2 Description of Domain

The topic under review is somewhat cross-discplinary relating to industrial assembly processes, business process modelling and learning systems. Overall the following subtopics are recognized:

1. Assembly process modelling and simulation

    (a) Fault modelling and simulation

2. Mathematical analysis of log data

    (a) Analysing interleaved process traces

    (b) Analysing delays and intervals

3. Process mining

4. Visualization of event logs

2.1 Keywords

The keywords relevant for the research were collected from a set of articles deemed especially relevant for this research. A representative set of articles was
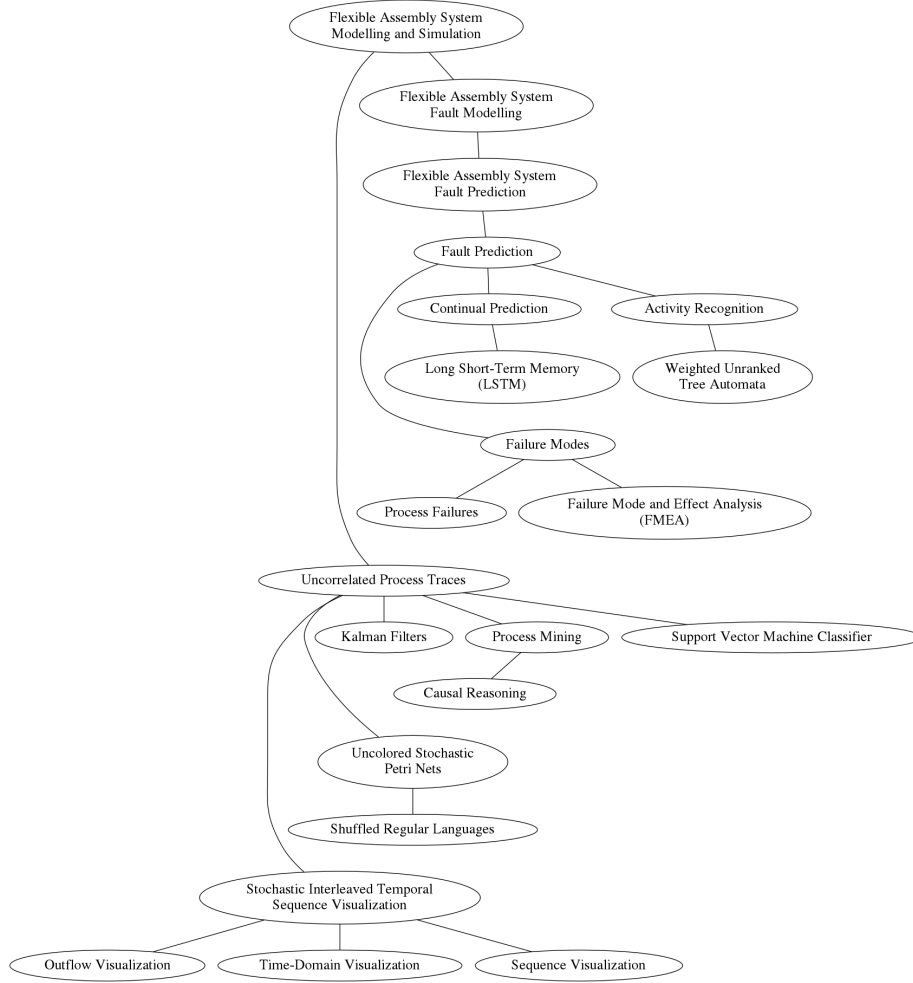
Figure 1: The mind map of the research topic

read and relevant keywords were picked from titles, abstracts and references. This set of keywords allows for directed browsing of relevant literature.

assembly process, assembly system, complex robotic system, compliant parts, dimensional quality, dimensional variation propagation, discrete event system, Failure mode and effects analysis (FMEA), failure rates, failure records, failure report, fault detection, fault mode, fixture variation, interleaving, leaf spring, multiple failure modes, Multi-Station Assembly Systems, outflow visualization, part variation, petri nets, Process algebra Petri Net (PPN), process control, process failures, process FMEA, reduction of irregularities, reliability, simulation, sonification, time-domain visualization, uncolored petri nets, uncorrelated event streams, variation propagation, visualization of sequences, welding gun variation

2.2 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?

3 Literature Review

Assembly systems, construction processes and business processes are often simulated using discrete event systems, for example in: [2, 3, 4, 5]. These simulations are used for process optimization[6], scheduling and assembly line design. The models of these systems are conveniently described using Process algebra Petri Nets (PPNs) [7]. Object Petri Nets have been used in modelling hierarchical systems[8]. For a lower abstraction level simulation IEC 61499 function block models[9] have been used for example in [10].

The literature about faults and deviances in assembly processes mainly focus on the faults in the assembled products. There are less academic publications about simulating process deviances in assembly processes such as unexpected process failures, human errors, and deadlocks. Cyberattacks have received more attention, for example in [8].

The model of the simulated system is created in production line design [11], by observing the system, or by interviewing the system specialists [12]. The model is changed for each batch of different products being produced.

The fault types relevant in the context of this research are faults in the assembly system itself, not faults in the products being assembled. The assembly system faults can be faults of specific assembly components, or systemic faults, such as deadlocks. Different kinds of faults relevant for the Flexible Assembly Systems have been described for example in [13].

A discrete event simulation output is a sequence of generated events, and the internal system state between the events. The sequence of generated events is useful as a corpus for learning algorithms if it reflects the real conditions and target phenomena well enough. In the case of anomaly detection, the simulation should be representative for the correct operation of the target system, and relevant process deviances should in principle be observable in the output data model.

Flexible Assembly Systems assemble a batch of products in parallel. Parallel processes create interleaved process traces. The set of all the interleaved process traces form a language. As this language consists of interleaved allowed sequences it is a shuffle language[14].

Process mining is the field of inferring the underlying business processes based on observed events and transitions. The process mining methods can be used to compare the supposed process model with the observed process model

to detect deviances. Process mining methods have not been previously used for Flexible Assembly System process modelling and verification.

The methods for describing process traces in the context of process mining are loosely based on the alpha algorithm[15] and as such, they expect process instances to be identified in the activity events to deinterleave the process traces and to infer causal relations of activities. "Event logs need to satisfy two main requirements: (a) events need to be ordered in time and (b) events need to be correlated (i.e., each event needs to refer to a particular case)."[16]

There is very little academic literature about process mining of uncorrelated event streams. Uncorrelated event streams are important modelling targets when the event sources are heterogeneous and not necessarily mapped to the specific steps in the known assembly process. This brings event sources such as emergency stop triggers, light curtains and smart gateways into the process information system even if the log messages they trigger are not correlated to specific process instances.

Multi-modal representation of the event streams for the purpose of evaluating the human-observable presentation of features in the data is relevant in context of this research. Having a human-detectable pattern in data where the same pattern is challenging to detect with automatic means highlights a potential avenue for additional research.

If a pattern is evident for humans in the data representation, it is in principle discoverable by learning algorithms.

Different human sensory systems are tuned for particular types of patterns, for example music and speech can be appreciated and recognized as audio, but the same signal patterns in visual representation are not recognized as well. A multi-modal representation utilizes different sensory-cognitive capabilities to better facilitate detection of embedded patterns in the data.

There is a lot of literature about discovering structure in sequences and time series by using different visual representations and projections [17, 18]. Visualizations are also used in model-based reporting[19]. Data representation as audio (Sonification [20]) has been investigated for example in [21, 22].

4 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 pub-

lished 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[23]. 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.

5   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[24] 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.

6   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.

7 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.[25]

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] M. Donath and R. J. GRAVES, "Flexible assembly systems: an approach for near real-time scheduling and routeing of multiple products," *The International Journal Of Production Research*, vol. 26, no. 12, pp. 1903–1919, 1988.

[2] V. Hlupic and S. Robinson, "Business process modelling and analysis using discrete-event simulation," in *Proceedings of the 30th conference on Winter simulation*. IEEE Computer Society Press, 1998, pp. 1363–1370.

[3] Y. Zhao, C.-B. Yan, Q. Zhao, N. Huang, J. Li, and X. Guan, "Efficient simulation method for general assembly systems with material handling based on aggregated event-scheduling," *Automation Science and Engineering, IEEE Transactions on*, vol. 7, no. 4, pp. 762–775, 2010.

[4] S.-K. Kang, H. Jung, I. H. Im, K.-Y. Chung, and J.-H. Lee, "Active discrete event simulation algorithm using probability distribution of shipbuilding process," in *Information Science and Applications (ICISA), 2013 International Conference on*. IEEE, 2013, pp. 1–3.

[5] A. H. A. Rahnama, "A fuzzy approach on simulating and optimizing the performance of a discrete-event production line," in *Computers and Industrial Engineering (CIE), 2010 40th International Conference on*. IEEE, 2010, pp. 1–5.

[6] N. Sadeghi and A. R. Fayek, "A framework for simulating industrial construction processes," in *Simulation Conference, 2008. WSC 2008. Winter.* IEEE, 2008, pp. 2396–2401.

[7] P. Falkman, B. Lennartson, and K. Andersson, "Specification of production systems using ppn and sequential operation charts," in *Automation Science and Engineering, 2007. CASE 2007. IEEE International Conference on.* IEEE, 2007, pp. 20–25.

[8] K. Wu, Y. Li, F. Chen, and L. Chen, "A method for describing industrial control system network attack using object petri net," *IEEJ Transactions on Electrical and Electronic Engineering*, 2015.

[9] I. TC65, "Wg6,"iec 61499: Function blocks for industrial-process measurement and control systems–parts 1 to 4," *Geneva: International Electrotechnical Commission (IEC)*, 2004.

[10] M. N. Rooker, T. Strasser, G. Ebenhofer, M. Hofmann, and R. V. Osuna, "Modeling flexible mechatronical based assembly systems through simulation support," in *Emerging Technologies and Factory Automation, 2008. ETFA 2008. IEEE International Conference on.* IEEE, 2008, pp. 452–455.

[11] H.-J. Bullinger and H. Sauer, "Planning and implementing a flexible assembly system supported by simulation," vol. 25. Taylor & Francis Ltd , Rankine Road, Basingstoke, Hants, RG24 8PR, UK, URL:http://www.tandf.co.uk, 1987, pp. 1625–1634, last updated - 2011-11-11. [Online]. Available: http://search.proquest.com/docview/24612706?accountid=11365

[12] J. A. B. Montevechi and J. D. Friend, "Using a soft systems methodology framework to guide the conceptual modeling process in discrete event simulation," in *Simulation Conference (WSC), Proceedings of the 2012 Winter.* IEEE, 2012, pp. 1–12.

[13] M. Cong, J. Zhang, and W. Qian, "Fault diagnosis system for automated assembly line," in *Intelligent Processing Systems, 1997. ICIPS'97. 1997 IEEE International Conference on*, vol. 2. IEEE, 1997, pp. 1478–1482.

[14] M. Berglund, H. Björklund, and J. Högberg, "Recognizing shuffled languages," in *Language and Automata Theory and Applications.* Springer, 2011, pp. 142–154.

[15] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 9, pp. 1128–1142, 2004.

[16] W. M. van der Aalst, "Process mining: discovering and improving spaghetti and lasagna processes," in *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on.* IEEE, 2011.

[17] T. Hein and O. Kramer, "Recognition and visualization of music sequences using self-organizing feature maps," in *KI 2010: Advances in Artificial Intelligence.* Springer, 2010, pp. 160–167.

[18] K. Misue, "Chronoview: A space-efficient method for visualizing temporal patterns," in *Computer Graphics, Imaging and Visualization (CGIV), 2014 11th International Conference on.* IEEE, 2014, pp. 1–4.

[19] M. Schuh, J. Sheppard, S. Strasser, R. Angryk, and C. Izurieta, "An ieee standards-based visualization tool for knowledge discovery in maintenance event sequences," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 28, no. 7, pp. 30–39, 2013.

[20] G. Kramer, *Auditory display*, ser. Santa Fe Institute studies in the sciences of complexity, Proceedings. Reading (Mass.): Addison-Wesley, 1994.

[21] E. S. Yeung, "Pattern recognition by audio representation of multivariate analytical data," *Analytical Chemistry*, vol. 52, no. 7, pp. 1120–1123, 1980.

[22] H. G. Kaper, S. Tipei, and E. Wiebel, "Data sonification and sound visualization," *Computing in science & engineering*, vol. 1, no. 4, pp. 48–58, 1999.

[23] 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.

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

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