

Artificial Data Generation of Complex Manufacturing Systems in Industry 4.0: Fulfilling Data Scarcity With a Generalizable Notation

Pedro Paulo Araújo de P. e Silva

Universidade Federal de Mato Grosso do Sul
Campo Grande - MS, Brazil
p-paulo@ufms.br

Paulo Victor Lopes

Aeronautics Institute of Technology
and Federal University of São Paulo
São José dos Campos - SP, Brazil
paulo.lopes@ga.ita.br

Filipe A. N. Verri

Computer Science Division
Aeronautics Institute of Technology
São José dos Campos - SP, Brazil
verri@ita.br

ABSTRACT

Complex manufacturing networks are composed by a multitude of interconnected processes, elements, and variables. Since, with the industry 4.0, these networks can generate a huge volume of data, so that existing bottlenecks, deadlocks and other dilemmas cannot be exposed easily. Consequently, advanced techniques such as process mining tools are needed in a way of enhancing this processes, by handling these data to provide the Knowledge Discovery in Databases (KDD).

Therefore, one of the advantages of the usage of process mining techniques is the optimization of a real process that can come through the Discrete Event Simulation (DES). These techniques allows not only a back-ward looking within the processes discovery and conformance checking, but a forward-looking by the what-if scenarios vision. Further, nowadays, manufacturing systems data is occasionally difficult to find, delaying the searches into the field. So, we propose artificial data as a key tool to work around with this problem.

Consequently, focusing to fulfill this lack of data, this project goal was to develop a universal simulation model that works with the well known Petri Net Markup Language (PNML) file format, together with the SimPy Python library that is an environment for DES. The advantages of the Petri Net notation is that this representation of real manufacturing systems has a mathematical foundation and a visual representation, enabling a more accessible computer-based approach.

In spite of that, this project consisted in three phases, the first one was the translation of a manufacturing network randomly generated made with a Poisson distribution, which comes in an acyclic directed graph representation (in which the edges represent network flow, and the nodes the machines), into a PNML file format. Secondly, the building of a simulation model in Python language, at this point, a metadata such as the buffer size and time-activity of each machine coming

from a JSON were added, concerning to have a more accurate simulation. And then, the third one was the generation of the event-logs in a way that process mining techniques such as process discovery and conformance checking would be possible.

Different networks were generated, through the variation of the numbers of machines and the number of production steps, and after the simulation, process discovery algorithms such as alpha miner, inductive miner and heuristic miner, were applied to the event-logs, in order to analyze the behavior of the simulation model.

Finally, this project enabled an easy way to simulate every process described in a petri net notation, with a high accuracy level and a close to reality scenario. All the codes are available in a github repository for further contributions.

KEYWORDS. Process Mining. Artificial Data. Petri Nets.