

Name : SAMAROO Sharmila, Master 2 MCI-Amérique Latine

Name of your level 1 : Angélique CHAGNEAU

Keywords specific to the content : Process Mining, BPM, Model Learning, Model Checking, System Under Learning (SUL), Generalization, DFA, Equivalence Query (EQ), L algorithm, TQs, MDL.

Source : e.g Googlescholar.com

[BDA/Business_process_analysis/L2_AissataDIAW/L1_AngéliqueCHAGNEAU/L0_SAMAROOSharmila/process_mining.pdf at main · BigData2024/BDA \(github.com\)](#)

Title- Article : Process mining meets model learning : Discovery deterministic finite state automata from event logs for business process analysis.

This article was written by Simone Agostinelli, Francesco Chiariello, Fabrizio Maggi, Andrea Marella, Fabio Patrizi and is available online on February 1st , 2023. Before getting to the heart of the matter, it is fundamental to understand the following processes :

Business process : *Is a set of activities executed in a given setting to achieve predefined business objectives.* Extract from the article

Model learning (ML) : *Refers to a group of test based.* Extract from the article.. It's divided in 2 classes : passive and active algorithms, but in this article we only study passive ML.

- Passive ML : *“performance decrease exponentially for logs with a significant activity.”* Extract from the article

This article deals with deterministic Finite State Automaton (DFA), which is an example of a learning model. DFA helps enterprises to predict future evolution of a business process and determine whether it is following the rules (below). To apply it, we need a model like (model learning) ML or MDL. It will help us to compare the effectiveness of DFA when using ML and MDL

First of all, we need to know the rules of DFA :

- Respect data awareness and constraints.
- Respect the alignments approach.

DFA's used for :

- Evaluate process conformity
- If the process violates or infringes these rules
- How to modify the template, to avoid breaking the rules

To be effective DFA's used Model Learning (ML), which function with 4 metrics. The result is given as a percentage.

Precision :	<i>Is used to assess the degree to which the behavior allowed by the process model are observed in the event logs. Cf.article</i>
Fitness:	<i>Quantifies the extent to which the discovered accurately reproduces the trace recorded in the log. C.f article</i>
Generalization:	<i>Estimate the extent to which a model deduced from a given journal will reproduce future behaviors not yet observed in a journal.</i>
Simplicity :	<i>Size of the model. C.f Article</i>

By applying this model we learn that when the value of ML result is close to 1% DFA model is reliable process. In additional, ML can be effective to complex systems. However, the limitations of this model are that it cannot handle noisy data and the L of ML is not suitable for application to “complex real life logs”, because DFA uses a lot of unreasonable time to generate an answer.

Furthermore they study another model which is MDL :

- Is a good comparison with DFA’s risky decision.
- Generates more understandable and simpler information than other algorithms.

Limitation :

- We learn that MDL: can generate simplest automata DFA compare to ML.

To conclude we saw the usefulness of using DFA into ML and MDL, we discover the advantages and disadvantages of these two models. In additional, to mitigate the noise in ML method, the solution could be to use a *noise” filtering mechanism before starting”* the ML. (extract from the article)