# Discovering Deterministic Finite State Automata from Event Logs for Business Process Analysis

Simone Agostinelli [1] Francesco Chiariello [1] Fabrizio Maria Maggi [2] Andrea Marrella [1] Fabio Patrizi [1]

[1]DIAG - Sapienza University of Rome [2]KRDB - Free University of Bozen-Bolzano

{agostinelli,chiariello,marrella}@diag.uniroma1.it

## Highlights

- Deterministic Finite State Automata (DFAs) are employed to perform formal reasoning tasks in Process Mining [1].
- We enable the automated discovery of DFAs from event logs.
- Novel process mining quality metrics tailored to DFAs and negative examples are introduced.

## Process Mining

- Process Mining [2] (PM): research area from Business Process Management.
- It analyzes process data recorded in event logs to gain insight into business processes.

## DFA for PM

- - Process Discovery [3],
  - Conformance Checking [4],
  - Compliance Monitoring [5].

## Model Learning

- Active Learning, e.g. L*.
- Passive Learning, e.g.
  - MDL, for positive examples,
  - RPNI and EDSM, for both positive and negative examples.

## Markovian Abstraction

A *k-th order Markovian Abstraction* [6] (*$M^k$-abstraction*, for short) over a set $\beta$ of traces is a finite graph $M^k = (S, E)$, with nodes $S$ and edges $E \subseteq S \times S$, such that:

- $S = \{-\} \cup S_1 \cup S_2 \cup S_3 \cup S_4$, where:
  - "$-$" is a special state;
  - $S_1 = \{\sigma \in \beta : |\sigma| \leq k\}$ is the set of traces of $\beta$ having length up to $k$;
  - $S_2 = \{\sigma[i,k] : \sigma \in \beta, i = 1, |\sigma| > k\}$, with $\sigma[i,k]$ denoting the $k$-length subtrace of $\sigma$ starting at position $i$, is the set of $k$-length prefixes of some trace in $\beta$ (with length greater than $k$);
  - $S_3 = \{\sigma[i,k] : \sigma \in \beta, |\sigma| > k, i = |\sigma| - k + 1\}$ is the set of $k$-length suffixes of some trace in $\beta$ (with length greater than $k$);
  - $S_4 = \{\sigma[i,k] : \sigma \in \beta, |\sigma| > k, 1 < i < |\sigma| - k + 1\}$ is the set of $k$-length subtraces of some trace in $\beta$ (with length greater than $k$), excluding prefixes and suffixes;
- $E = \{(-, \sigma) : \sigma \in S_1 \cup S_2\} \cup \{(\sigma, -) : \sigma \in S_1 \cup S_3\} \cup \{(\sigma, \sigma') : \sigma, \sigma' \in S_2 \cup S_3 \cup S_4, \exists \hat{\sigma}, i \text{ s.t. } \hat{\sigma} \in \beta, |\hat{\sigma}| > k, 1 \leq i \leq |\hat{\sigma}| - k, \sigma = \hat{\sigma}[i,k], \sigma' = \hat{\sigma}[i+1,k]\}$.

## Precision

Given a log $\ell$ and a DFA $G_\mathcal{M}$, let $M^k_\ell = (S_\ell, E_\ell)$ and $M^k_{G_\mathcal{M}} = (S_{G_\mathcal{M}}, E_{G_\mathcal{M}})$ be their respective M$^k$-abstractions, $C$ the Levenshtein-distance-based cost matrix, and let $\mu_C : E_{G_\mathcal{M}} \to E_\ell$ be a partial function, solution of the assignment problem represented by $C$. The (Markovian-abstraction-based) k-th order precision of $G_\mathcal{M}$ wrt $\ell$ is defined as:

$$MAP^k(\ell, G_\mathcal{M}) = 1 - \frac{\Sigma_{e \in E_{G_\mathcal{M}}} C(e, \mu_C(e))}{|E_{G_\mathcal{M}}|},$$

by taking $C(e, \mu_C(e)) = 1$, if $\mu_C(e)$ is undefined.

## Fitness

Given a log $\ell$ and a DFA $G_\mathcal{M}$, let $M^k_\ell = (S_\ell, E_\ell)$ and $M^k_{G_\mathcal{M}} = (S_{G_\mathcal{M}}, E_{G_\mathcal{M}})$ be their respective M$^k$-abstractions, $C$ the Boolean cost matrix, and let $\mu_C : E_\ell \to E_{G_\mathcal{M}}$ be a partial function, solution of the assignment problem represented by $C$. The (Markovian-abstraction-based) k-th order fitness of $G_\mathcal{M}$ wrt $\ell$ is defined as:

$$MAF^k(\ell, G_\mathcal{M}) = 1 - \frac{\Sigma_{e \in E_\ell} C(e, \mu_C(e)) F_e}{\Sigma_{e \in E_\ell} F_e},$$

where $F_e$ stands for the frequency of edge $e$ in $E_\ell$ and taking $C(e, \mu_C(e)) = 1$, if $\mu_C(e)$ is undefined.

## Conclusions

- Active learning algorithms are not suitable to generate DFAs from real-life event logs.
- Declare Miner and passive learning algorithms construct DFAs with similar values of generalization and precision.
- Passive learning algorithms generate simpler DFAs than Declare Miner.

## Future Work

- Learn LTL$_f$ formulae:
  - directly from logs, or,
  - going through Alternating Finite Automa.
- both approaches possible with SAT or ASP techniques.

## References

[1] Simone Agostinelli, Francesco Chiariello, Fabrizio Maria Maggi, Andrea Marrella, and Fabio Patrizi. Process mining meets model learning: Discovering deterministic finite state automata from event logs for business process analysis. *Inf. Syst.*, 114:102180, 2023.

[2] Wil M. P. van der Aalst. *Process Mining - Data Science in Action, Second Edition*. Springer, 2016.

[3] Fabrizio Maria Maggi, R. P. Jagadeesh Chandra Bose, and Wil M. P. van der Aalst. Efficient discovery of understandable declarative process models from event logs. In *24th International Conference on Advanced Information Systems Engineering (CAiSE 2012)*, pages 270–285, 2012.

[4] Massimiliano de Leoni and W. M. P. van der Aalst. Aligning event logs and process models for multi-perspective conformance checking: An approach based on integer linear programming. In *International Conference on Business Process Management*, pages 113–129. Springer Berlin Heidelberg, 2013.

[5] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil M. P. van der Aalst. A framework for the systematic comparison and evaluation of compliance monitoring approaches. In *17th International Conference on Enterprise Distributed Object Computing (EDOC 2013)*, pages 7–16. IEEE, 2013.

[6] Adriano Augusto, Abel Armas-Cervantes, Raffaele Conforti, Marlon Dumas, and Marcello La Rosa. Measuring fitness and precision of automatically discovered process models: A principled and scalable approach. *IEEE Trans. Knowl. Data Eng.*, 34(4):1870–1888, 2022.

## Acknowledgements

We thank the H2020 project DataCloud (Grant number 101016835), the Sapienza grant BPbots, the UNIBZ project CAT, the ERC Advanced Grant WhiteMech (No.755834228) and the EU ICT-48 2020 project TAILOR (No. 952215).

TAILOR
Foundation of Trustworthy AI:
Integrating Learning, Optimisation and Reasoning