Master Thesis Proposal

Probabilistic process mining

Supervisors:	Carlos Cotrini (ccarlos@inf.ethz.ch)
Professor:	Prof. David Basin
Issue Date:	May 31, 2019
Submission Date:	TBD

1 Introduction

Process mining offers organizations the possibility to better understand how internal processes work [1]. Process mining consists of algorithms that analyze logs produced by the organization's systems to discover process models that illustrate how products and services are created. Those models are aimed at giving insights for optimizing those processes and, therefore, reducing costs and increasing productivity.

Process mining is gaining traction in industry. Gartner published a market guide for process mining in 2018 that included several common use cases for process mining and an overview of current vendors [2]. However, to our knowledge, process mining uses only combinatorial algorithms and we see great potential for algorithms based on machine learning and artificial intelligence.

In this thesis, we approach the field of process mining from a machine-learning perspective. We shall develop a *probabilistic process miner* that, in contrast to the current state of the art, produces probability distributions over process models rather than producing actual models. Moreover, we shall develop model validation techniques for process mining.

2 Objective

This thesis aims to investigate the following:

1. Maximum-likelihood approaches for process mining.

2. Bayesian approaches for process mining.

3. Deterministic annealing approaches for process mining.

4. Model validation techniques for process mining.

3 Tasks

- Become familiar with the current state of the art in process mining, e.g. [3][4][5].

- Propose and justify a model validation technique for process mining.

- Propose a maximum-likelihood approach, a Bayesian approach, and/or a deterministic annealing approach for process mining.

- (Optional) Propose a universal approach for mining processes, DCR graphs, and automata.

- Conduct experiments on realistic process mining scenarios.

4 Deliverables

- At the end of the second week, a detailed time schedule of the project must be given and discussed with the supervisor.

- At the end of the project a presentation of 30 minutes must be given during an Infsec group seminar. It should give an overview as well as the most important details of the work.

- Software and configuration scripts must be delivered to the supervisors.

- A final report consisting of an introduction, a discussion on the related work, presentation of main contributions, and experimental results. Three copies of this report must be delivered to the supervisor.

References

[1] Thomas H. Davenport and Andrew Spanyi. *What Process Mining Is, and Why Companies Should Do It*. `https://hbr.org/2019/04/what-process-mining-is-and-why-companies-should-do-it`. Harvard Business Review, 2019.

[2] Marc Kerremans. *Market Guide for Process Mining*. `https://www.gartner.com/en/documents/3870291`. Gartner, 2018.

[3] Wil Van Der Aalst. *Process mining: discovery, conformance and enhancement of business processes*. Vol. 2. Heidelberg: Springer, 2011.

[4] Alessandro Berti, Sebastiaan J. van Zelst, and Wil van der Aalst. *Process Mining for Python (PM4Py): Bridging the Gap Between Process-and Data Science.* arXiv preprint arXiv:1905.06169 (2019).

[5] Maikel Leemans, Wil MP Van Der Aalst, and Mark GJ Van Den Brand. *Hierarchical performance analysis for process mining.* Proceedings of the 2018 International Conference on Software and System Process. ACM, 2018.