

*Public summary of the PhD thesis of Niek Tax
PhD defense date: 19th of June 2019*

Process Mining techniques for unstructured event data

Process mining techniques aim to extract actionable insights from event data. Traditionally applications of process mining have focused mainly on the business process management (BPM) domain, where the event data that is generated is often structured as an effect of normative process specifications that are in place during execution of a business process.

This thesis focuses on process mining on event data that does not originate from a business process and where such a normative specification is absent. A typical property of such event data is that it is way less structured than the event logs that are found in the BPM domain, which causes many challenges to existing process mining algorithms. We propose a set of algorithms and techniques to analyze and understand event data that is highly unstructured, putting special emphasis on analysing event data from smart homes that record daily activities of human life.

In this thesis, we propose a set of techniques to analyze such data, which can be grouped into two categories of techniques.

The **first** category of methods focuses on preprocessing event logs in order to enable process discovery techniques to extract insights from unstructured event data. In this category we have developed the following techniques:

- An unsupervised machine learning approach to refine event labels based on the time at which the event took place, allowing for example to distinguish recorded eating events into breakfast, lunch, and dinner.
- An approach to detect and filter from event logs so-called chaotic activities, which are activities that cause process discovery methods to overgeneralize.
- A supervised machine learning approach to abstract low-level events into more high-level events, where we show that there are situations where process discovery approaches overgeneralize on the low-level event data, but are able to find precise models on the high-level event data.

The **second** category focuses on mining local process models. This involves cases where there might not be a global process model that explains all of the behavior in the event data, there can still exist several frequent patterns that each capture a local fragment of behavior.

Several techniques are introduced in the area of local process model mining, including a basic method, fast but approximate heuristic methods, and constraint-based techniques. All techniques have been implemented and evaluated on real-life data from smart home experiments. However, the techniques are more generally applicable in other domains of unstructured event data.

Title of PhD thesis: Mining Insights from Weakly-structured Event Data. Promotor: Prof. Wil M.P. van der Aalst (TU/e and RWTH Aachen University). Other main parties involved: Philips Research.