Topic 1: Visual Analytics in Process Mining

Process mining is a research field that focuses on analyzing business processes that are executed on IT systems, based on their behavior as recorded in an event log. With the growing availability of event log data and the advanced maturity of tools and algorithms, process mining has recently also gained widespread industry adoption, with many organizations implementing tools or conducting projects as a central part of their business process management strategy. This means that process mining results are presented to business analysts and managers, who have little expertise in process analytics, but have to make decisions based on their results. Visualizing these results in a user-centric way is therefore a vital part of establishing process mining in organizations.

Visual analytics is a research field that focuses on facilitating analytical reasoning by means of data analytics techniques and interactive visual interfaces. It combines multiple disciplines including computer science, cognitive psychology, and graphic design in order to visualize complex situations and assists humans in understanding and advancing them.

The goal of this seminar thesis is to identify current usage of visualization in decision making. The participant should therefore conduct a structured literature review in the field of visual analytics. It should be identified what is visualized how and for which purpose. The elements and frameworks used for that should then be applied to process mining. The participant should show how the frameworks used in visual analytics can be embedded into the process mining context. Specifically, the applications to conformance checking are of interest.

Topic 2: Causal Analysis in Process Mining

Process mining denote a family of techniques for the analysis of event logs that record the behavior of business processes, as captured from enterprise information systems. Those techniques are typically explorative, i.e., they can identify or describe correlations in the event log data but cannot establish or reason about the causes of the identified patterns. Such causal relationships, however, are more valuable than mere correlations, because process managers can be sure that changing their process according to a causal insight will improve their process. Recently, process mining researchers have therefore started to adopt techniques from causal machine learning to detect causal relationships in event log data.

The goal of this seminar thesis is to identify the current state of the art regarding the application of causal analysis techniques in process mining. Therefore, the participant should conduct a structured literature review to identify all recent papers that apply a causal machine learning approach to event log data. Those papers should be analyzed regarding (1) the use case that they support (e.g., process prediction or discovery), (2) the perspective that they take (e.g., control-flow), and (3) the causal analysis technique they employ. Because the literature in this area is suspected to be sparse, the participant is expected to analyze each identified paper in detail and ideally detect future research potentials in this area.

Topic 3: Process Discovery with the Inductive Miner

Process mining discovery techniques aim to generate a model of how a process is conducted from the log data of process executions generated by information systems supporting the process. There exist many process discovery algorithms used in different application domains, one of which is the Inductive miner (IM) algorithm. The Inductive miner discovers the process tree for the log recursively with a "divide and conquer" strategy of dividing the log, building part of the process tree, then working on each division separately. Several extensions to the IM exist, such as the IM-Infrequent and the inductive visual miner.

The goal of this seminar thesis is to systematically explore the different extensions of the IM. Accordingly, the participant is expected to conduct a structured literature review to describe the initial IM algorithm and compare it to the IM extensions. This includes 1) explaining each extension and how it differs from the original IM, 2) explaining the reason(s) behind the extensions, 3) describing when each extension is suitable to be used, and 4) evaluating the quality of each extension (e.g., in terms of fitness, precision, generalization, soundness, etc.).

Topic 4: Hierarchical Process Discovery in Task Mining

Process discovery is perhaps the most commonly used process mining technique. Its goal is to derive a process model from an event log that analysts can then use to understand and improve the execution of processes in an organization. Various process discovery algorithms exist, ranging from the simple alpha miner to the more powerful heuristic and inductive miners. However, all the commonly used discovery techniques discover flat process models that do not incorporate event log information beyond the control-flow of activities. This can be insufficient in the context of task mining or robotic process automation, where process models discovered from detailed User Interaction Logs are often overly complex and lack important contextual information. One possible solution to this is to discover more expressive, hierarchical process models with hierarchical process discovery.

In this seminar thesis, the participant should first conduct a structured literature review to identify different hierarchical process discovery algorithms. The algorithms should then be analyzed with regard to how they function and which parameters and a-priori information they depend on. Finally, the algorithms should be evaluated with respect to their applicability in task mining.

Topic 5: Data-Driven To-Be Modeling

Organizations operate using a set of business processes that are designed according to best practices. These processes need to be captured and modeled for different purposes, the most important of which is documenting and improving organizational operations. However, process modeling is not an easy task, and some organizations operate without having models for their processes. To-Be models (or normative models) model process rules and guidelines as assigned and directed by organizations. There exist approaches for developing such To-Be models, the most recent of which is data-driven To-Be modeling from process execution logs.

The goal of this seminar thesis is to identify the current state of the art regarding data-driven To-Be modeling. The participant should conduct a structured literature review listing existing approaches for data-driven To-Be modeling. The SLR should focus on explaining 1) the approach with its steps and technical details, 2) how the approach is evaluated, and 3) the limitations and future directions for improving the approach.

Topic 6: Explaining Deviations in Conformance Checking

Conformance checking is one of the main tasks of process mining. It comprises techniques for checking the relation between a designed process model and the real-life behavior of a process, with the goal of identifying and analyzing deviations between them. State-of-the-art conformance checking approaches have become computationally very efficient and robust to many kinds of process behavior. However, most of those approaches focus only on the control-flow perspective, i.e., only take the activities from the event log into account. This might limit their applicability in many practical applications, where the order of the activities might not be the (sole) relevant factor for the conformance of a process. Instead, other process attributes like cost or time might play a major role.

As conformance checking itself cannot be considered as a means to an end, further analyses can prove useful in order to explain why deviations from the designed process model occur. Users will be enabled to improve the process conformance by that.

The goal of this seminar thesis is to identify approaches that explain differences in conformance (i.e., in fitness). Therefore, the participant should conduct a structured literature review regarding the (domain independent) support of deviation analysis. In a second step, the identified approaches should be analyzed with regard to their application potential to conformance checking (i.e., why the conformance differs in an event log).

Topic 7: User Interaction Logs in other Domains

User interaction (UI) logs are high-resolution event logs that record low-level, manual activities performed by a user during the execution of a task in an information system. Each event in a UI log corresponds to a single interaction between the user and the software application, such as clicking a button. In a process mining context, UI logs are typically used to automate process executions by means of Robotic Process Automation, but they can also be used for other purposes, such as the identification of usability issues in software.

However, the idea of using system-generated logs of user interactions is not unique to the process mining domain. Other disciplines, such as web analytics or human-computer interaction, have also been using UI logs to analyze user behavior in different contexts. The goal of this seminar thesis is to systematically explore the usage of UI logs in disciplines other than process mining. Therefore, the participant should conduct a structured literature review in order to identify the disciplines in which UI logs are used. Next, the participant should characterize the structure of a typical UI log per discipline and compare it to the UI logs that are typically used in a process mining context, identifying commonalities and differences, for example with regard to attributes and granularity.

Topic 8: Relevant Setup Choices for Next-Activity Prediction

Predicting the next activity in a process trace has become one of the main applications of supervised machine learning in process mining, and a plethora of different prediction techniques have been proposed. When evaluating these techniques, researchers use very different configurations, data sets, and parameters. For example, the setups may differ in train/test split strategies, number of cross-validation folds, filtering of infrequent variants, and trace length padding approach. As a result, it is difficult and often misleading to compare the performance figures reported for next-activity prediction algorithms. This is a known problem in the process mining community, but so far there has not been a detailed analysis of all the setup choices used in the evaluation experiments.

The goal of this seminar thesis is to review the literature on next-activity prediction and to compile a comprehensive overview of setup choices made by different authors. The review should also briefly evaluate the identified setup choices with respect to how they may affect the reported results.