CHAPTER 1

RELATED WORK

In this chapter, the researches related to the work presented in this thesis are presented. Firstly, latest trends and studies in the process mining area are explained. Then researches from cross-organizational process mining, which is the main idea behind this research, is introduced. Following these, studies related to similarity in process mining is mentioned. Finally, performance and conformance analysis approaches in process mining area are presented. In each section, survey of related studies are tried to be constrained with the subject of this thesis.

1.1 State of the Art in Process Mining

In this section, important milestones in process mining and current research trends will be presented. Studies in process mining area have roots based on process discovery techniques for software engineering. These original techniques are studied to discover workflows using neural networks, purely algorithms and Markovian approaches [14]. However, using process mining in the workflow management area is introduced by Agrawal et al. and this study tries to find workflow graphs given event logs and identify edge conditions between nodes [3]. In the following researches, many different approaches based on hierarchical structuring, dependency and frequency graphs; and heuristic methods are suggested to address the same problem. While some of these algorithms are focused on creating partial solutions, some of the algorithms are used as a foundation for further expansions like "Alpha Algorithm" of van der Aalst et al.[11]. Spread of process mining is not only in kept in academia,
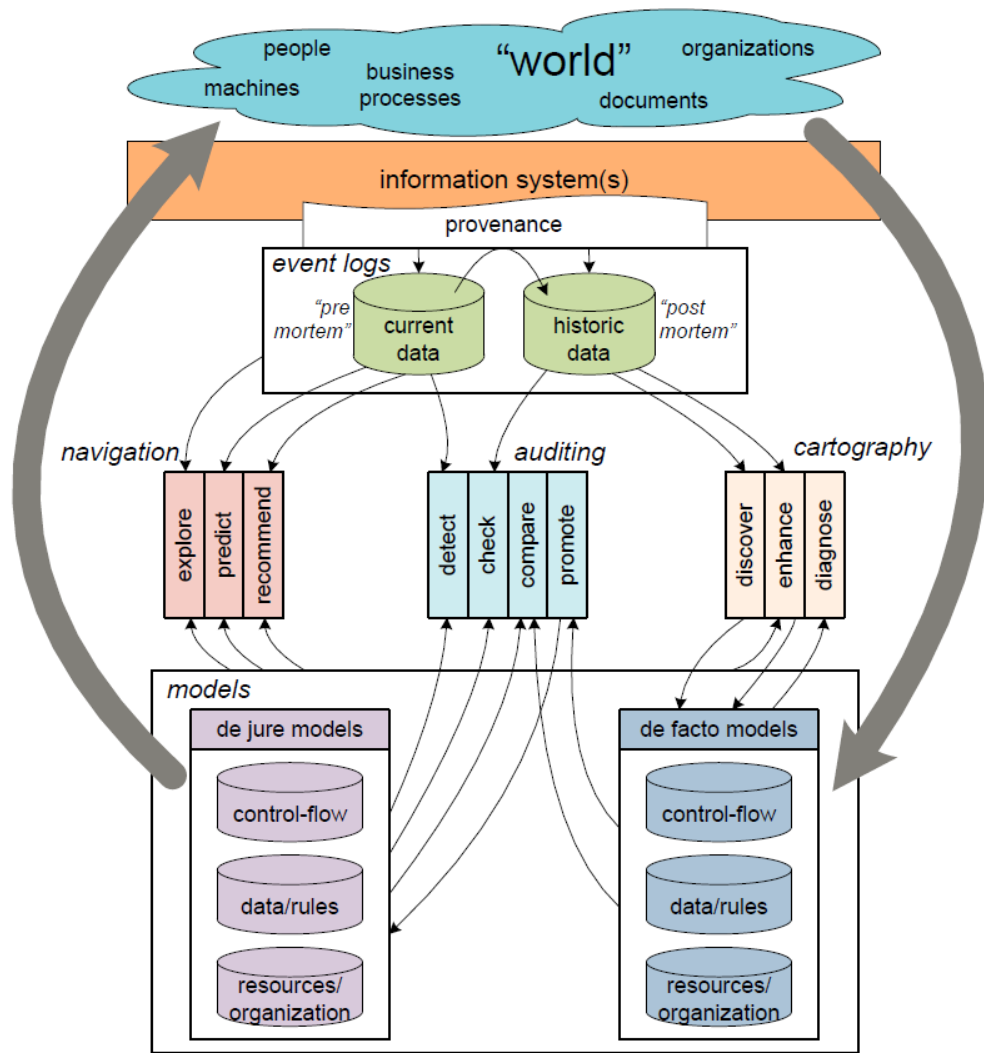
Figure 1.1: Process mining framework by van der Aalst

but also many vendors presented tools and software to discover processes and help organizations manage their workflows [1].

Current challenges in the process mining area can be presented best mentioning spectrum and its relations. As shown in the Figure 1.1 from the researches of van der Aalst [15], process mining spectrum is extensive and highly inter-related. Starting with the provenance, process mining deals with gathering and keeping event logs as *current data* and *historic data*. In this level, *post mortem data* means the logs and information gathered from completed events whereas *pre mortem data* means information related to ongoing cases. This separation helps us to use historical knowledge to exploit and gain advantage over current business operations. At the very bottom of the diagram,

models are presented as *de facto models* and *de jure models*. *De jure models* are created with the aim of describing reality as it should be whereas *de facto models* aims to describe reality as it is. Within each model category a mixture of perspectives, in other words point of views from different information levels, can be mixed. In order to fill the gap between event logs and models, ten different process mining activities are listed under three categories [9]:.

Cartography: In this category the main aim is to create maps of the real world activities by using process models.

> **Discovery:** Discovery is focused on extracting process models from event logs.

> **Enhancing:** Enhancing activities are based on improving *de facto models* with the information hidden in the event logs.

> **Diagnose:** Diagnose activities stand for identifying the problems caused by directly process models.

Auditing: In this category, activities related to confronting the model and the reality are collected.

> **Detect:** Detection is based on comparing the *de jure models* with the ongoing processes to generate alerts if shifts occur in the reality.

> **Check:** Checking is defined to identify deviations occurred in the past.

> **Compare:** Comparison of *de facto models* and *de jure models* shows the difference between planned and the reality.

> **Promote:** Promotion includes partially updating *de jure models* from the information gathered from *de facto models*.

Navigation: In this category, process mining methods are used like a navigation application.

> **Explore:** Event data and models are used in combination to explore business processes.

> **Predict:** Prediction activities combine the information from past events and make predictions about ongoing processes.

Recommend: Suitable actions can be suggested by using the prediction and contextual information.

Within this framework, the study presented in this thesis is a combination of discovery from cartography, promoting from auditing and recommendation from navigation. In other words and within the metaphor of spectrum, this study tries to discover maps using the discovery techniques and promotes the partially best locations in the map to create recommendations for travelers in their navigation applications.

1.2 Cross-organizational Process Mining

In this section, related researches and trends in cross-organizational process mining will be presented. Cross-organizational mining is based on cross-correlation of workflows and the realized activities in different organizations. The main challenge of this topic in process mining is that comparing processes and their performances of different organizational units in an objective approach. This objective approach is open to be enhanced with the process context, namely the environment of the process that is executed. Most of the studies in this area are currently studying to reveal the possible opportunities and some initial approaches to address main challenges.

In the study of Bujis et al. [5], they indicated the importance related to increase of Software-as-a-Service (SaaS) and cloud computing infrastructure usages. As a result of this increase, more and more organizations will use a Shared Business Process Management Infrastructure (SBPMI) and it is an opportunity for different organizational units to learn from each other in such infrastructure. The approach presented in this study is based on three questions and three metric groups to answer these questions. *a*) Which organizations support my behavior with better process models? *b*) Which organizations have better behavior which my process model supports? and *c*) Which set of organizations can I support with my process model? While answering these questions, they used *simplification based metrics* to indicate better process models; and *throughput time metrics* to indicate better behaviors. In this study, it is shown how a generic framework can be used to highlight the main idea behind cross-organizational process mining.

In the study of van der Aalst [12], using configurable process models is proposed for the organizations sharing the same infrastructure and doing the similar work. Configurable process models are defined as a family of process models where each organization can use this family with their configurations according to their business needs. This approach not only creates behaviors needed by each organization but also creates a basis to compare and learn within process mining framework. The configurable process models are formalized by "Causal Nets", which is a notational language based on input and output bindings of each node. Although this study does not answer to all challenges related to cross-organizational-process-mining, a formalism of configurable processes models is presented to address learning and conformance checking.

Cross-organizational process mining is divided into intra-organizational and inter-organizational with two basic ideas: *collaboration* and *exploiting* commonality [13]. In his study, van der Aalst defined collaboration for distributed work between multiple organizations and commonality as ability of using same process models and infrastructure between the organizations. In order to exploit these two ideas, two partitioning dimensions are suggested in [13]:

Vertical Partitioning: Process instances, namely cases, distributed over several organizations which collaborate to complete a complex activity.

Horizontal Partitioning: Process parts, namely tasks or activities, are shared within organizations like jigsaw puzzle parts.

For these orthogonal dimensions, a number of questions and challenges are presented in the study [13]. For the vertical partitioning, where organizations are aiming to share infrastructure and knowledge to learn from each other, it is mentioned that supervised learning methods like classification can be used. Notion of this thesis is based on vertical partitioning of cross-organizational process mining with the idea of unsupervised learning where predictor variables related to performances of organizations are used.

1.3 Similarity in Process Mining

In this section, prominent researches related to similarity in process mining will be presented with their results. With the emerging attention in business processes, organizations become aware of the fact that they can exploit the business processes and their similarities [4]. In addition, most of the large organizations have repository of process models of similar business operations [6]. There are three main different approaches proposed to this solution in the literature currently.

First approach, proposed by Dijkman et al. [6], is based on similarity metrics as *a*) node matching similarity; *b*) structural similarity; and *c*) behavioural similarity. Result of the study [6] indicates that using these three similarities can differentiate comparable process models and within these metrics structural similarity is the most prominent one.

Second approach by Bujis and Reijis is an analytical approach [4] to compare the process models of different organizations that does similar works. Proposed algorithm is based on creating an alignment matrix between observed and realized models. This inter-relation is also used to compare different variants of the same process by different organizations. In addition, they suggested their method as a framework to further standardize a process of common interest.

The third and final approach is based on frequently occurring mismatches between similar business processes [7]. In this study [7], a set of mismatch patterns are derived from the different department of the same organization. Although this research does not present a complete set of possible mismatch patterns, the provided set is comprehensive to identify similarities and differences of the processes of same operations. In this thesis, combination of metric and mismatch pattern approaches are used to identify variations between process models of different organizations.

1.4 Performance and Conformance Analysis in Process Mining

In this section summary of studies related to performance and conformance analysis will be presented. Event logs do not only contain task sequences but also contain

time, resource and contextual information. Analysis of process models with these additional information is used within performance analysis framework to highlight bottlenecks or make predictions. However, in order to undertake a performance analysis there is a need of replaying reality (event logs) on the expectation (process models) and check the conformance of reality to plan [10].

In the study conducted by Rozinat and van der Aalst [8], a conformance framework is proposed by two metrics *fitness metrics* and *appropriateness metrics*.In this study [8], fitness metric is based on replaying the event log on the process model and counting the number of missing or remaining tokens. In other words, replaying and conformance of event logs over process models is modelled as a token passing formalism. On the other hand, appropriateness metrics are based on how accurate the process model in describing reality within a degree of clarity. Simplicity, precision and generalization attributes of the process models are taken into account while calculating appropriateness metrics.

Token passing approach to conformance has a major drawback when the process model and reality of event logs do not fit completely. In this case, there are over-estimated process models that are too general, in other words with too much behavior other than the reality. In order to overcome this drawback, heuristic and optimization based methods are proposed by other researchers. In this thesis study, an extended version of optimization based approach presented in [2] is used to replay logs on the process models. Since the main goal of this study is not to evaluate conformance of logs and process models, the method presented by Adriansyah et al. is extended to calculate performance indicators while replaying the logs.

REFERENCES

[1] Rafael Accorsi, Ernesto Damiani, and Wil van der Aalst. Unleashing operational process mining. 2014.

[2] Arya Adriansyah, Boudewijn F van Dongen, and Wil MP van der Aalst. Conformance checking using cost-based fitness analysis. In *Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International*, pages 55–64. IEEE, 2011.

[3] Rakesh Agrawal, Dimitrios Gunopulos, and Frank Leymann. *Mining process models from workflow logs*. Springer, 1998.

[4] Joos CAM Buijs and Hajo A Reijers. Comparing business process variants using models and event logs. In *Enterprise, Business-Process and Information Systems Modeling*, pages 154–168. Springer, 2014.

[5] Joos CAM Buijs, Boudewijn F van Dongen, and Wil MP van der Aalst. Towards cross-organizational process mining in collections of process models and their executions. In *Business Process Management Workshops*, pages 2–13. Springer, 2012.

[6] Remco Dijkman, Marlon Dumas, Boudewijn Van Dongen, Reina Käärik, and Jan Mendling. Similarity of business process models: Metrics and evaluation. *Information Systems*, 36(2):498–516, 2011.

[7] Remco Matthijs Dijkman. *Mismatch Patterns in Similar Business Processes*. Beta, Research School for Operations Management and Logistics, 2007.

[8] Anne Rozinat and Wil MP van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.

[9] Wil Van Der Aalst. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media, 2011.

[10] Wil Van der Aalst, Arya Adriansyah, and Boudewijn van Dongen. Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2):182–192, 2012.

[11] Wil Van der Aalst, Ton Weijters, and Laura Maruster. Workflow mining: Discovering process models from event logs. *Knowledge and Data Engineering, IEEE Transactions on*, 16(9):1128–1142, 2004.

[12] Wil MP van der Aalst. Business process configuration in the cloud: how to support and analyze multi-tenant processes? In *Web Services (ECOWS), 2011 Ninth IEEE European Conference on*, pages 3–10. IEEE, 2011.

[13] Wil MP van der Aalst. Intra-and inter-organizational process mining: Discovering processes within and between organizations. In *The Practice of Enterprise Modeling*, pages 1–11. Springer, 2011.

[14] Wil MP Van der Aalst and AJMM Weijters. Process mining: a research agenda. *Computers in industry*, 53(3):231–244, 2004.

[15] WMP van der Aalst. Challenges in business process mining. *Applied Stochastic Models in Business and Industry (to appear)*, 2010.