UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Lekso Migriauli

A Systematic Literature Review On Process Discovery Aspects Of Process Mining Applied In Industrial Contexts

Master's Thesis (30 ECTS)

Supervisors:
Fabrizio Maria Maggi
Chiara Di Francescomarino

Tartu 2021

A Systematic Literature Review On Process Discovery Aspects Of Process Mining Applied In Industrial Contexts

Abstract:

Process mining is a discipline, part of Business Process Management (BPM) that involves analyzing business processes on the basis of event logs. Due to its potential benefits, process mining has gained significant academic interest and hundreds of papers have been published on the topic. In the scope of this paper, a Systematic Literature Review (SLR) was conducted to ease the burden on business analysts by providing a clear, value-driven framework for process mining use cases and providing an overview of papers that contribute to the value of transparency, more specifically business process discovery. A total of 839 papers were scanned and filtered throughout the SLR. 267 of these papers were identified to be about process mining methods that contribute to the value of transparency and 183 of them were about business process discovery. These 183 papers are analyzed in greater detail throughout the thesis.

Keywords:

Process Mining, Business Process Management, Systematic Literature Review, Business Process Discovery

CERCS: P170

Süstemaatiline kirjanduse ülevaade tööstuse kontekstis rakendatava protsessi kaevandamise protsesside avastamise aspektidest

Lühikokkuvõte:

Protsessi kaevandamine on eriala, osa äriprotsesside juhtimisest (BPM), mis hõlmab äriprotsesside analüüsimist sündmuste logide põhjal. Protsessikaevandamine on oma võimalike eeliste tõttu pälvinud märkimisväärse akadeemilise huvi ja sellel teemal on avaldatud sadu töid. Selle artikli raames viidi läbi süstemaatiline kirjanduse ülevaade (SLR), et leevendada ärianalüütikute koormust, pakkudes selge, väärtuspõhise raamistiku protsesside kaevandamise kasutamise juhtumitele ja andes ülevaate paberitest, mis aitavad kaasa läbipaistvuse väärtusele , täpsemalt äriprotsesside avastamine. Skaneeriti ja filtreeriti kogu peegelkaameras kokku 839 paberit. Leiti, et neist dokumentidest 267 käsitlesid protsesside kaevandamise meetodeid, mis suurendavad läbipaistvuse väärtust, ja 183 neist käsitles äriprotsesside avastamist. Neid 183 artiklit analüüsitakse kogu lõputöö jooksul üksikasjalikumalt.

Võtmesõnad:

Protsessi kaevandamine, äriprotsesside juhtimine, süsteemne kirjanduse ülevaade, äriprotsesside avastamine

CERCS: P170

Table of Contents

1 Introduction

Business processes are what companies do whenever they deliver a service or a product to customers (Dumas et al, 2013). Any combination of activities such as modeling, automation, execution, optimization of business processes done in order to support enterprise goals is known as Business Process Management (BPM) (Palmer, n.d.).

In the modern age of computers and software, business processes are controlled and supported by online systems. These systems log each of the individual activities executed as a part of business processes. The end result of such logging is an event log. Event logs consist of traces, which capture a single instance of a business process implementation, also known as a case. Each trace, in turn, contains individual, timestamped events or activities from the business process.

Process mining is a discipline, part of Business Process Management (BPM) that involves analyzing business processes on the basis of such event logs. Using process mining techniques, it is possible for organizations to capture and analyze the actual execution and performance of their processes. Due to its potential benefits, process mining has gained significant academic interest. Hundreds of papers have been published on the topic, making it harder for business analysts to identify (1) the prominent use cases of process mining methods and (2) what business-oriented questions such methods can address.

This paper aims to answer the following research questions:

> RQ1: “What are the main use cases for existing process mining methods?”
>
> RQ2: “What business-oriented questions do existing process mining methods address?”

By answering these questions, the final goal is to ease the burden on business analysts by providing more information about what questions each of the process mining methods addresses and helping them find the most suitable process mining method for their use cases.

In order to reach the thesis goals, a systematic literature review was conducted according to the guidelines presented by Kitchenham (2004). Academic publications related to process mining methods and techniques were gathered from electronic databases using search strings. The initial total of 2293 papers was filtered out to a final of 839 papers. A Value-Driven framework for process mining use cases was developed that maps out the business values that BPM can help organizations achieve. This paper mainly focuses on analyzing 183 out of 267 papers that focus on the value of transparency in terms of business process model discovery.

The remainder of this paper is organized as follows: section 2 will present the Systematic Literature Review protocol; section 3 will introduce the Value-Driven BPM framework for process mining use cases, while section 4 will describe SLR results in better detail, focusing on the value of transparency. Finally, Section 5 concludes the thesis.

2 Systematic Literature Review Protocol

This section provides an overview of the SLR protocol used and iterates on the main research questions that the SLR answers.

This paper aims to bring more clarity to the process mining methods, what enterprise values they help to achieve and what business-oriented questions do they address. To that end, two research questions (RQ) were defined. The first research question (RQ1) aims to identify and categorize possible use cases of process mining methods and it is defined as: "What are the main use cases for existing process mining methods?". The second research question (RQ2) aims to elicit specific business-oriented questions that the identified process mining methods address. RQ2 is defined as follows: "What business-oriented questions do existing process mining techniques address?"

In order to answer these research questions, an SLR was conducted. SLR follows the guidelines for a literature review defined by Kitchenham (2004) and Webster and Watson (2002). With the SLR, existing research work was collected and analyzed. The SLR consists of planning, execution, and reporting (Kirchenham, 2004). This section will include the planning and execution phases of the conducted SLR.

In order to find relevant papers, SLR was divided into two parts. The first part involves searching for similar, SLR-type papers on specific process mining methods. Such papers include, for example, SLR on Process Discovery conducted by Augusto, Conforti et al. (2018). Similarly, SLRs on Conformance Checking (Dunzer, Stierle, et al., 2019) and Predictive Monitoring (Di Francescomarino, Ghidini, et al., 2018). Each of these SLRs contains a list of final papers, which have been extracted for the current SLR. The second part of the SLR involves the papers that have applied process mining methods to real-life event logs.

The papers from two parts of the SLR were merged together and content screening was applied to them. Both parts of the SLR follow the same guidelines as proposed by Kitchenham (2004) and Webster and Watson (2002). The steps include developing search strings, identifying search sources, filtering the results according to predefined criteria, identifying additional relevant papers through backward referencing, and extracting data according to a predefined form.

2.1 Planning

This section provides a more detailed description of the planning and execution steps conducted during the SLR.

The planning phase of the SLR mainly includes search string development, search sources identification, selection criteria, and data extraction criteria definition. The search strings were derived based on the research questions and the scope of the study, as directed by Kitchenham (2004) and Kitchenham, Mendes, et al. (2007).

2.1.1 Search Strings

Since the first part (SLR review) of the SLR aims to gather SLR-type work on process mining, the search string applied was ("process mining" AND "systematic literature review"). However, several known process mining use cases, such as concept drift, have not been systematically reviewed. Furthermore, SLR may not cover all publications. Therefore, the search strings for the second part (PM review) were derived from the research questions. The term "process mining" was derived from the scope of the SLR. In addition, "workflow mining” was also included in the search string as this term is often used interchangeably for “process mining”. “real-life”, “real-world” and “case study” were also included in the search string based on the scope of this SLR, since the goal is to find process mining techniques that are applicable to real-life event logs. Based on the above terms, the following search strings were formulated.

- ST1: (“process mining” AND “real-life”)
- ST2: (“process mining” AND “real-world”)
- ST3: (“process mining” AND “case study”)
- ST4: (“workflow mining” AND “real-life”)
- ST5: (“workflow mining” AND “real-world”)
- ST6: (“workflow mining” AND “case study”)

2.1.2 Search Sources – Electronic Databases

The next step for the SLR planning is to define where the papers will be searched for. That is to say, defining specific search sources and electronic databases where the papers will be extracted from. As per Levy and Ellis (2006) and Bereton, Kitchenham et al. (2007), search sources were restricted to electronic databases in the field of computer science. As a result, Scopus and Web of Science were the chosen databases as they include information on where the most research on process mining is published.

2.1.3 Selection Criteria

This section defines the selection criteria for the SLR. By defining exclusion criteria (EC) and Inclusion Criteria (IC), it is possible to make paper selection reproducible. These criteria allow for filtering all the search result papers in order to end up with a more relevant and fruitful list. The following subsections define inclusion and exclusion criteria according to the guidelines suggested by Fink (2019).

2.1.3.1 Exclusion and Inclusion Criteria for SLR Review

In order to filter out the initial list of discovered papers, exclusion and inclusion criteria were defined. The first exclusion criterion (EC1) aims to remove duplicates. As suggested by Kofod-Petersen (2012), both exact duplicates - papers with the same title from the same authors appearing in different digital libraries - and version duplicates - papers from the same authors with approximately the same title - were removed. From the version duplicates, the most recent work was kept.

The next exclusion criteria aim to filter out the papers that are not understandable or accessible. To that end, the second exclusion criterion (EC2) aims to remove papers that are not in English and the third exclusion criterion (EC3) aims to eliminate papers that are not available in digital libraries or have extra restricted access (cannot be accessed by the University subscription).

As for the inclusion criteria, the first inclusion criterion makes sure that the found work is actually focused on process mining. For example, one of the search hits was a paper by

Badakhshan et al., (2019), which is an SLR about Business Process Management. While the paper mentions the concept of process mining, it is not the primary focus.

The last inclusion criterion (IC2) is about keeping only the papers that conducted an SLR on a specific process mining use case. For example, papers such as the ones about the evaluation of process mining algorithms by Naderifar et al., (2019) were excluded.

To summarize, the defined criteria for the first part, the SLR review, are as follows.

- EC1: Is the SLR study a duplicate?
- EC2: Is the SLR study in English?
- EC3: Is the SLR study digitally accessible?
- IC1: Is the SLR study on process mining?
- IC2: Does the SLR study specify process mining use case?

2.1.3.2 Exclusion and Inclusion Criteria for PM review

The exclusion criteria that apply for the first part (SLR review) are the same for the second part (PM review). In addition, the papers that are less than 5 pages were excluded, since they do not contain enough information for analysis. This is the last exclusion criterion (EC4).

As a result, the defined exclusion criteria for the second part, the PM review, are as follows.

- EC1: Is the paper a duplicate?
- EC2: Is the paper in English?
- EC3: Is the paper digitally accessible?
- EC4: Is the study less than 5 pages?

Next, the inclusion criteria were defined in order to only keep relevant work. The papers that are not within the scope of process mining, such as papers about the mining of minerals were excluded (IC1).

The second inclusion criterion (IC2) aims to only keep the papers that apply process mining to real-life event logs. In addition to being applied to real-life logs, it is important that the discussed process mining methods are valuable and applicable to industry (business) needs (IC3). This criterion excludes papers that discuss process mining applications that are not related to business, such as comparing the capabilities of discovery algorithms to manage noisy event logs (Nuritha, Mahendrawathi, 2019). The last inclusion criterion (IC4) aims to only keep the papers that describe business case setting in enough detail so as to elicit the business questions being addressed.

The resulting list of the inclusion criteria for the PM review is as follows.

- IC1: Is the study within the domain of process mining?
- IC2: Does the study apply the process mining technique to a real-life event log?
- IC3: Does the study use the process mining technique to address a business need?

- IC4: Does the study sufficiently describe the case setting so as to elicit the questions being addressed?

2.1.4 Data Extraction

The last step in planning the SLR is data extraction. This step aims to define an objective form of data extraction in order to reduce the bias opportunity (Kitchenham, 2004).

The data extraction form was developed according to the suggestions by Fink (2019) and Okoli (2015). The data extraction step consists of two parts - extracting paper metadata and extracting data about process mining use case and the questions being addressed by a process mining technique. The first, metadata extraction part involves extracting paper ID, title, authors, and publication year.

2.2 Execution

This section describes the execution of the planning steps described in section 4.1. Searches were conducted in February 2020.

Application of the search string for the first part (SLR review) resulted in 132 hits from the Scopus database and 61 hits from Web of Science. Hence, the total number of candidate papers for the SLR review is 193. With the exclusion criteria, only 60 papers were kept. The first inclusion criterion (IC1) resulted in removing another 15 papers, while IC2 removed an additional 26 papers. This process finally resulted in 19 remaining papers. By applying backward and forward referencing, 2 additional papers were discovered, taking a total list of SLR works to 21.

Data extraction for the SLR review involved extracting all the studies included in the final list of these 21 SLR publications. These results were later merged with the PM review papers. A total of 702 papers were extracted from the 21 SLR publications.

For the PM review (second part), applying the search strings resulted in 1021 hits from the Scopus database and 570 hits from Web of Science. This total of 1591 hits was combined with the 702 papers from the SLR review, bringing the full list to 2293 papers. the initial phase of cleaning filtered 91 papers out. With EC1, 889 duplicates were found and removed and another 109 papers were removed for EC4 (study less than 5 pages). This resulted in a total of 1204 papers. IC1 filtered out another 80 publications, while IC2, IC3, and IC4 removed a total of 316 papers. This process ended up in a total of 808 papers. Applying forward and backward referencing identified an additional 21 papers, bringing the total number of papers to 839. Out of these 839, 267 papers fell under transparency, out of which 183 papers about process model discovery were summarized in this work.

Figure 1 shows the number of papers included in each category relevant to the scope of this work.

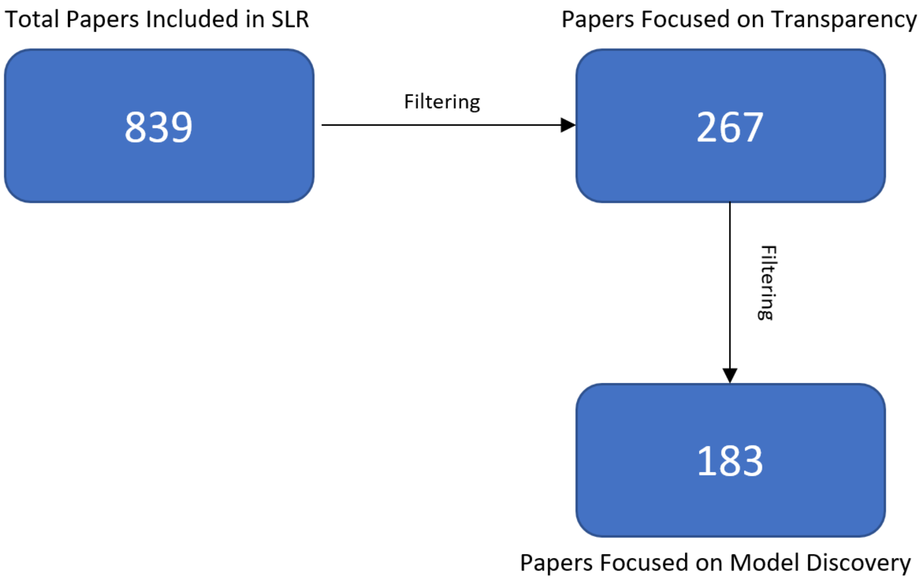

Figure 1. Paper Selection Process

Figure 2 shows the distribution of resulting papers included in SLR by the year of their publishing.

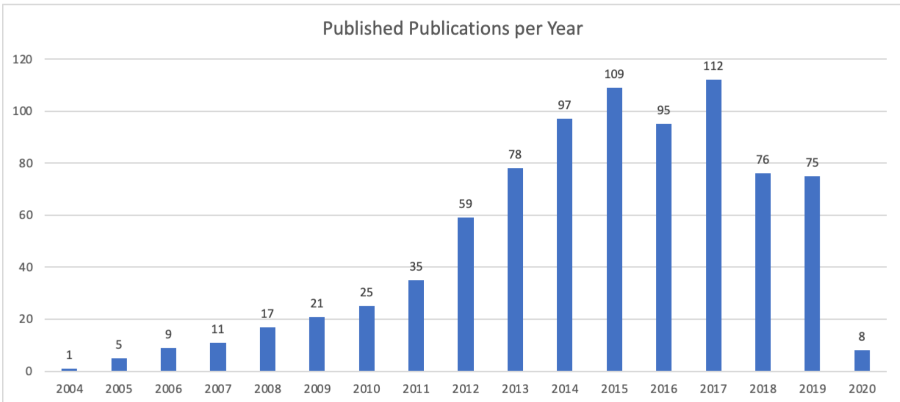

Figure 2. Publications Distribution Over Years

The complete list of selected papers is provided as supplemental material and can be accessed from a public link[1]. This section has provided an overview of the SLR protocol. The results of the SLR will be discussed in the following sections.

[1] https://drive.google.com/file/d/12lm47aEo3QLtz-loBOOY4yyfQR_p8MAT/view?usp=sharing

3 Value-Driven Framework

This section will introduce a value-driven framework for process mining use cases. The overall structure and description of the framework, as well as its threats to validity, are presented. In addition, the main use cases of process mining methods and the questions such methods address are discussed.

Value-Driven BPM (VBPM) maps out six main corporate values that organizations engaged in BPM try to achieve. These six values are efficiency, quality, compliance, agility, integration, or networking. At the core of all of these values lies transparency. Transparency here relates to the clarity of process executions inside organizations. “Only an organization that has a shared understanding of its processes can start reflecting on better ways to design and operate them” (Franz and Kirchmer, 2013). Hence, transparency is essential for VBPM.

As already mentioned, organizations engage in BPM in order to achieve one or many values out of efficiency, quality, compliance, agility, integration, or networking. Organizations concerned with increasing efficiency focus on their business processes and optimizing them in order to achieve efficiency gains. One of the specific examples for this could be finding and eliminating waste in the process. On the other hand, some organizations might choose quality over efficiency. Such companies utilize BPM techniques to link process executions to the final output (quality) of the product or service.

Similarly, organizations that are subject to strict regulatory requirements have prioritized the value of compliance. Such organizations aim to standardize their processes so that there are fewer drifts from the main process flow and hence, fewer compliance problems. One of the examples of such organizations could be a pharmaceutical company. Such companies have very strict regulatory rules on the processes, as well as the end product of their processes. They can gain a lot of value from process standardizations. On the other hand, some companies value agility more. They gain more value by being flexible and changing the process flow as the business requirements change. One example of such an organization could be a seaside hotel that has a boom in demand during summer but experiences a decrease during winter months. Agility is the core value that keeps such businesses profitable.

The remaining two values are networking and integration. Integration involves integrating internal resources such as employees into the planning and design of business processes, while networking focuses on the external partners and resources for the same goal (Franz and Kirchmer, 2013).

Since VBPM is based on interviews and surveys with actual organizations, it best captures the values that such companies aim to achieve via BPM. Since the goal of this paper is to categorize process mining use cases in order to help business analysts and organizations in achieving their enterprise goals, VBPM is a fair starting point for the proposed framework. Also, VBPM allows keeping the focus on actual enterprise values that BPM helps to achieve rather than the specific phases of BPM lifecycle or process mining methodologies.

The following subsection will describe in detail the proposed framework and how it builds upon the VBPM framework described above.

3.1 Structure of the Framework

The proposed value-driven framework for process mining use cases builds upon the aforementioned VBPM framework in order to best categorize process mining use cases. The proposed framework consists of two parts to clearly align with the research questions. Hence, the first part of the framework is concerned with eliciting and categorizing the main use cases of process mining, while the second part includes specific questions that the existing process mining methods address. The first part (categorization of main use cases) is based on the value-driven management of business processes introduced by Franz and Kirchmer (2013).

Since the proposed framework is based on the VBPM framework, it uses the same enterprise values to categorize the main process mining use cases. Two values, integration and networking are excluded from the framework since process mining cannot significantly benefit any of these values. As already mentioned, integration is about involving and integrating internal resources such as employees into the process planning and design, while networking does the same with external resources. While process mining can be useful in discovering and designing business models, the involvement and accessibility to these models are out of the scope.

The proposed framework groups process mining use cases into quality, agility, conformance, efficiency, and transparency values.

Transparency includes process mining methods such as Discovery, Model Repair, Model Enhancement, Social Mining Network, Goal Modelling, and Rule Mining. All of these methods enhance transparency over business processes. For example, process Discovery is all about discovering or building a process model based on event logs. Model Repair involves improving on the discovered process models to better reflect the actual execution of the process. Model Enhancement means adding data to the process models. Social Mining Network draws the link between the resources of a business process. Goal Modelling or Goal Mining describes the relationships between process goals. Rule Mining is about defining the business rules that affect the path that the business process takes.

Process mining methods under the value of efficiency include methods that analyze process execution data and draw conclusions about the performance of the processes. It also includes methods such as Process Performance Optimization which helps draw insights on how to improve and optimize process performance and efficiency.

Process mining methods that help enhance quality include Variant Analysis and Deviance Mining. Variant Analysis involves comparing different variants of the same process to each other, while Deviance Analysis allows identifying reasons for deviations during the process execution.

Two process mining methods are included in the compliance category. These methods are Conformance Checking and Compliance Monitoring. Conformance Checking involves comparing an actual process execution with a predefined process model. Compliance Monitoring is an ongoing process of monitoring process executions that are non-compliant with predefined rules and constraints.

Lastly, process mining methods that help enhance agility include Predictive Monitoring, Concept Drift, and Perspective. Predictive Monitoring is about trying to predict outcomes of ongoing process executions. Concept Drift describes how process executions differ over time. Perspective involves suggesting specific actions to take for ongoing process execution in order to achieve desired outcomes.

Figure 3 provides a high-level view of the proposed framework. As described, the process mining methods are categorized into values of transparency, agility, efficiency, compliance, and quality. Each value contains a list of methods that help enhance the value. For example, Variant Analysis helps enhance quality for an organization. Then, in alignment with research questions, the questions that these process mining methods address are defined along with brief descriptions and references to the papers. For example, for Social Mining Network, the question defined is "what are the relationships between the resources of a business process?"

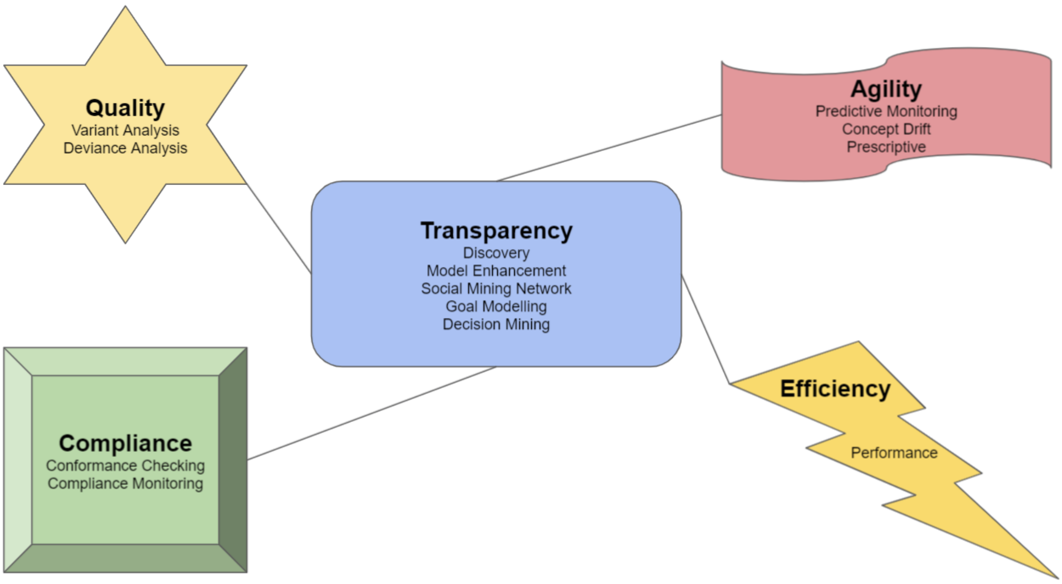

Figure 3 Value-Driven Framework For Process Mining Use Cases

The questions that process mining methods under a transparency address are usually descriptive from the business process model perspective (Discovery). They also try to describe relationships between resources or process goals (Social Mining Network and Goal Modelling) or improve upon discovered process models (Model Enhancement and Model Repair).

Process mining methods under efficiency quantitatively describe actual process behavior. In addition, these methods can also address comparative questions, as well as the cause of the differences (e.g. Deviance Mining).

Similarly, process mining methods under compliance address questions that describe and compare business process executions to each other, and are also able to detect where, how, and why the actual behavior differs from the initially designed process.

Methods under agility also address descriptive questions, but in some cases (Prescriptive) can also improve upon descriptions and comparisons in order to come up with a recommended action.

3.2 Threats to Validity

This section describes the potential threats to validity of a proposed framework for categorizing process mining methods.

Since the proposed framework is elicited based on a Systematic Literature Review (SLR), typical SLR inaccuracy threats apply. These might include bias during the filtering of papers, inaccuracies during data extraction, etc.

In order to avoid such threats as much as possible, the SLR was conducted based on the guidelines described by Kitchenham (2004) and Webster and Watson (2002). In particular, well-known electronic databases were used for finding the relevant work, and backward referencing was conducted in order to avoid the inaccurate exclusion of relevant papers. In addition, the SLR protocol is presented in order to ensure replicability. It must be noted that, due to the updates to the search algorithms or databases, possible deviations in results are expected when trying to replicate the protocol.

4 Systematic Literature Review Results

Process model discovery is one of the process mining use cases that falls under the *transparency* value. Since the process models can be either procedural, declarative, or the hybrid of the two, business-oriented questions that this use case answers have been divided into three parts.

This section provides an overview of how different works have answered the business-oriented question "How are the instances of a business process executed?" Different aspects related to this question are discussed in the following subsections.

4.1 Procedural Model Discovery

Procedural process models are a type of process models that aim to precisely describe the control flow of business processes. In procedural process models, all activities included in a process are present and organised so as to describe the process flow best.

Different aspects of procedural process discovery are presented in the following subsections.

4.1.1 Artifact-centric

Artifact-centric process models aim to describe complex processes as a collection of interacting artifacts (van Eck, Sidorova, and van der Aalst, 2017).

In (van Eck, Sidorova, and van der Aalst, 2017) the focus is on how discovered artifacts are related to each other. Moreover, the paper provides the ways of visualization and qualification of the interactions between the discovered artifact-centric models. The approach was implemented as a ProM plug-in.

Enterprise resource planning (ERP) systems are often used in business processes management to manage business documents and record event data. Lu, Nagelkerke, Wiel, and Fahland (2015) present the discovery of artifact-centric process models from ERP systems. Specifically, the work elicits the life cycle of business objects and their interaction. The approach was validated as a part of the case studies.

Artifact-centric process model discovery brings notable benefits. However, many process mining methods cannot be used directly, as they serve monolithic modeling. Popova, Fahland, and Dumas (2015) present a method for automated discovery of artifact-centric process models. The approach interprets the problem so that it becomes possible to use an already existing wide range of (non-artifact-centric) process model discovery methods.

According to Popova and Dumas (2014), synchronization conditions arise in artifact-centric processes, consisting of groups of interacting artifacts, each with its life cycle. In the model discovery process, methods are designed to find synchronization conditions over a finite, size-fixed set of events. Although discovery methods fail to handle the challenge of obtaining synchronization conditions when a set of events has variable size. Popova, & Dumas (2014) propose a method to discover such synchronization conditions from event logs.

Eck, Sidorova, and Van Der Aalst. (2018, June) consider that very often the main focus of artifact-centric process discovery is on the representation of the artifacts. Therefore the paper's objective is to provide a strategy to discover interactions between correlated behaviors in artifacts and analyze their performance. The solution has been evaluated using real-life data and has been implemented as a ProM plug-in.

4.1.2 Business Process Model and Notation (BPMN)

Procedural business process models can be represented in several different ways. BPMN is the standard notation for representing business process models. Discovering process models from event logs is one problem but representing the discovered models in a standard and understandable way is another. This subsection will present the works that have tackled the problem of process model discovery in the scope of BPMN.

The tradeoff in discovering process models from event logs is between the model accuracy and the structural complexity (Augusto, Conforti, Dumas, La Rosa and Bruno, 2018). The approach presented in (Augusto, Conforti, Dumas, La Rosa and Bruno, 2018) states to discover accurate process models first and then structure them. After conducting an experimental evaluation on real-life and synthetic logs, the discover-and-structure approach proved itself to be effective from an accuracy and structural complexity perspective.

Kalenkova, Burattin, Leoni, Aalst, and Sperduti (2018) introduce an approach for discovering multi-perspective hierarchical BPMN models. The paper integrates various process mining techniques. The objective of the work is to fill the gap between consolidated process mining techniques and a wide range of BPMN-compliant tools.

For automated discovery of BPMN models, the technique, BPMN miner, is presented in (Conforti, Dumas, García-Bañuelos, and La Rosa 2016). Compared to existing model discovery techniques, the BPMN miner is able to detect and filter out noise in the event log. Thus models become more accurate and simpler. The validation of the proposed technique was done on synthetic logs.

Augusto, Conforti, Dumas, La Rosa, and Polyvyanyy (2018) describe an automated process discovery method, Split Miner, that discovers BPMN models from event logs. The paper claims that the technique has a faster execution time compared to other modern discovery methods. The models discovered by the Split miner have low complexity of branches and high fitness and precision. It is worth noting that Split Miner produces deadlock-free process models with concurrency if it is not restricted to building block-structured process models.

BPMN miner 2.0, a technique for discovering BPMN models, is presented by Conforti, Augusto, La Rosa, Dumas, and Garcia-Banuelos (2016). The method takes a log in XES format as input and discovers the BPMN process model. BPMN miner 2.0 is available as a standalone Java tool and a ProM and an Apromore plugin. The paper would be interesting for process mining researchers as well as for those who are involved in process mining employing BPMN.

The process discovery technique discussed by Wang, Wen, Yan, Sun, and Wang (2015) works on event logs that do not contain attributes such as primary key and foreign key. The technique builds BPMN models with sub-processes and multi-instance markers with event logs containing fewer event attributes. The experimental evaluation of the proposed method was done on synthetic and real-life logs.

The framework proposed in (Al-Ali, Cuzzocrea, Damiani, Mizouni, and Tello 2019) states that there is no one-to-one mapping between process model activities and event logs, instead, they are on different levels. The method uses machine learning to map low-level event logs to high-level activities. The framework was extended with a BPMN translator. The approach was validated with case studies with real-life data.

4.1.3 Patterns

In addition to discovering specific process models from event logs, process mining can be useful to discover patterns from available logs. This process is known as pattern mining and works related to it are presented in this subsection.

Leemans, and Van Der Aalst (2015) focus on pattern mining. The paper states that for a wide range of process mining techniques process instances play a crucial role. On the other hand, pattern mining does not favor it. The work presents an approach that discovers a recurrent group of events from event logs and considers the fact that events are related to process instances. The solution has been evaluated using real-life data and the ProM plug-in has been developed.

In the business processes, one can find broad repeated activities, patterns. Instead of detecting such patterns manually, the approach presented in (Ferreira, and Thom, 2012) seeks to make this process automatic. For discovering patterns the method uses semantic annotations of the event logs. The proposed solution was tested as a part of the case study.

The streaming process discovery algorithm, namely StrProM, is presented by Hassani, Siccha, Richter, and Seidl (2015). StrProM discovers sequential patterns from the stream by building prefix trees. The method continuously updates trees. The trees are used for generating the model. The approach was evaluated using real-world data and showed promising results from a quality and performance perspective.

Li, Bose, and Van Der Aalst (2011) propose an approach that improves model discovery by considering context-dependent patterns selected by users and creating a hierarchical process map. The patterns are used to fetch event data at a higher abstraction level. The method has been tested with real-life logs and integrated with ProM.

Deeva, and Weerdt (2019) propose a local pattern mining approach to deal with the challenge of expressing complex and unstructured processes with an understandable process model. The purpose of the technique is to enlarge specific parts of data and make local pattern detection in the area. Moreover, there is a discussion about how to use the method to avoid expensive tasks, such as discovering all available patterns and how to deal with unstructured data.

4.1.4 Online

A large number of process modeling methods are working on the uninterrupted data. However, when events are coming at a high rate, the model requires continuous updates. Process model discovery technique that is able to handle real time, continuous stream of event logs is known as online process discovery.

A large number of process modeling methods are working on the uninterrupted data. However, when events are coming at a high rate, the model requires continuous updates. Thus, the approach for discovering the business process model online is presented by Leno, Armas-Cervantes, Dumas, La Rosa, and Maggi (2018). Compared to other online discovery methods, which exploit a large amount of memory to maintain a model's clarity, the approach uses cache memory management to deal with the memory constraint and maintain accurate results.

Redlich, Molka, Gilani, Blair, and Rashid (2014) state that business processes are changing rapidly in contemporary organizations, which produce hundreds of events per second. The approach discussed in Redlich, Molka, Gilani, Blair, & Rashid (2014) proposes different modifications for the Constructs Competition Miner (CCM), which will enable the method to discover a run-time process model from data streams.

Evolving organizations require their systems to be able to adapt quickly to changing environments. Burattin, Sperduti, and Van Der Aalst (2012) propose a framework that can modify the Heuristics Miner to use the algorithm for model discovery from streams. The evaluation was done on real-life and synthetic logs. The Stream-aware Heuristics Miner was implemented in ProM.

Burattin, Sperduti, and Van Der Aalst (2014) focus on control-flow discovery from a stream of event data. The paper describes the ways of adapting Heuristics Miner to runtime discovery. The approach was tested on real and artificial logs.

Discovering process models from event streams, as (Zelst, Dongen, and Van Der Aalst, 2017) states, brings challenges and opportunities. One of the challenges is how to handle unlimited data with restricted memory. The paper proposes a generic architecture for lifting process discovery to the streaming domain. The approach was evaluated and implemented in ProM.

Examination of learning processes and providing knowledge of the processes online is Okoye, Tawil, Naeem, and Lamine (2016)'s primary target. For improved model discovery, data is extracted from learning execution environments and is transformed into an executable format for mining. The study describes a semantic process mining approach that enriches event data from learning processes. The method is able to predict individual learning patterns and outcomes through further semantic analysis of the discovered models.

In the companies where process mining activities are recurrent (e.g., twice a month), event logs are extracted from information systems by process mining tools, and processes are analyzed. For discovering a model, extracted data should go through a preprocessing step. The challenge is that growing data increases the data preprocessing time. In (Syamsiyah, Dongen, and Van Der Aalst, 2018) discussion is about the environments where event data can be extracted in real-time, and every generated log is preprocessed. As a result, data would be ready for analysts to take further steps. The approach was integrated with ProM and was tested on real-life logs.

4.1.5 Error handling

Real-life event logs often do not include explicit references that either the process was cancelled or an error occurred during execution. In order for process discovery to be precise, such errors and cancellation regions should be properly handled.

(Leemans, and Van Der Aalst, 2017) and (Kalenkova, and Lomazova, 2014) focus on cancellation features and propose a process discovery method that examines cancellation regions from event logs. For example, if a bank loan request was canceled, the system needs to handle such cases and produce appropriate logs. The methods discussed in the papers are oriented to equip the model with cancelation features. The approach presented in Leemans & Van Der Aalst (2017) was evaluated using real-life logs and was implemented as a ProM plugin.

4.1.6 Sub-processes

Large and complex business processes can be made more understandable by utilizing subprocesses. Subprocesses are easier to model and understand manually but automating their discovery is a challenge.

Nguyen, Dumas, Hofstede, La Rosa, and Maggi (2019) discuss an automated stage-based discovery method from event logs to avoid the discovered process model being flat. The method first identifies process stages from event logs and then discovers the model. The approach is compared with a basic divide-and-conquer technique. The divide-and-conquer technique, when applied to real-life logs, produces complex models. The method described in the report outperforms the divide-and-conquer approach concerning model accuracy and complexity.

Identifying potential bottlenecks in the business processes and getting performance analysis from the discovered model is Dongen, and Adriansyah (2010)'s primary focus. A clustering algorithm was proposed for discovering the model from event data. The model discovered by the method has nodes that are in many-to-many relationships with activities. Finally, performance information was projected onto the derived model. The approach has been validated on real-life logs.

Conforti, Dumas, García-Bañuelos, and La Rosa (2016) present the technique called BPMN miner for automated discovery of BPMN models. The method identifies a process-subprocess hierarchy from the event data. Models are discovered according to the derived order of processes(parent, subprocess). The BPMN miner can detect and filter out noise from event data. As a result, models become more accurate and more straightforward. The technique was validated on synthetic logs.

According to Nguyen, Dumas, ter Hofstede, La Rosa, and Maggi (2017, June), the primary challenges for process mining techniques are clarity and scalability. A common approach is to decompose processes into stages to mine each set of stages separately. However, when the process mining methods are confronted with real-world data, decomposed stages are far from the desired results. The paper proposes a method that does the decomposition of processes with high modularity. An evaluation of real-life event logs showed that the approach is close to manual decomposition that analysts would produce manually.

Nguyen, La Rosa, Dumas-Menijvar, and ter Hofstede (2017) discuss an approach based on business process stages. Stages can be considered as sub-processes. The paper states that dividing processes into stages can improve existing process mining techniques.

The hierarchical division of business processes presented in Molka, Redlich, Drobek, Zeng, and Gilani (2015) can help build more domain-specific models preferred by analysts. The paper focuses on discovering Business Process Model Notation (BPMN) models and on evolution strategies to mine domain models. The approach contributes to evolutionary algorithms, which proves itself to be a practical alternative way to real-world cases.

Conforti, Dumas, García-Bañuelos, and La Rosa (2014) claim that existing techniques generally build flat process models to which the concept of sub-process is not familiar. The approach presented in discovers BPMN models that contain sub-processes. The method analyzes the relation between the data attributes associated with events in order to define sub-processes. The validation was done on synthetic and real-life logs. It showed that models discovered by the method are more accurate and more understandable than the ones derived from flat process discovery methods.

The primary goal of the discussion presented by Ekanayake, Dumas, García-Bañuelos, and La Rosa (2013) is to improve the understanding of the discovered process model. The paper states that the collection of models produced with clustering strategies may be accompanied by excess complexity. The proposed approach splits the models into variants and sub-processes in order to reduce complexity. The technique allows users to put a complexity limit on the model. The approach was tested on real-life logs, and the result showed that the discovered models were reduced in size compared to those discovered by other trace clustering techniques.

Sun, and Bauer (2016) aim to solve the problem related to the discovered models that suffer from inaccuracy and complexity. The proposed approach optimizes the quality of the process model based on the clustering technique. The method builds high-quality abstraction models by considering the quality of discovered sub-models. In order to validate the applicability of the technique, a case study was carried out.

Process mining techniques that discover flat process models are struggling in the cases when processes contain sub-processes. Weber, Farshchi, Mendling, and Schneider (2015) aim to solve the problem of overgeneralization and lack of fitness in the mined models. The paper introduces the method for hierarchical discovery of processes. The capability of conformance checking is also added to the proposed method. The approach was evaluated with real-life event data.

The process discovery technique discussed by Wang, Wen, Yan, Sun, and Wang (2015) targets discovering BPMN models from event logs with fewer event attributes, such as case id, task name, start time, and end time. The technique analyzes the hierarchy of events and introduces an algorithm to discover sub-processes and multi-instances. The final required process model is composed of the models discovered from sub-processes. Synthetic and real-life logs were used for the experimental evaluation.

According to Jalali (2014), process models record different processes, which consider various concerns, such as privacy, authorization, and auditing. The aspects are spread in the models and are mixed up with processes. As stated in Jalali (2014), cross-cutting concerns are the concerns that present tangling and scattering relation to process models. The report discusses how models with cross-cutting concerns are discovered from event data. The evaluation of the proposed solution was done on a case study.

Yzquierdo-Herrera, Silverio-Castro, and Lazo-Cortés (2013) focus on trace clustering and grouping sub-processes that form the analyzed process. The paper claims that process diagnosis at early stages provides a comprehensive view of the processes. By diagnosis, the report implies performance analysis, pattern mining, and anomaly detection in the processes. The proposed approach can discover and group sub-processes, and it offers an accurate analysis of processes.

4.1.7 Context discovery

Since process discovery is meant to make process insights available for business analysts, it is important that the discovered models contain all the information that analysts require. In addition to conventional process models, analysts find it important to see relevant contexts during process execution.

According to Štajner, Mladenić, and Grobelnik (2010) context is a group of particular needed information. The work is about discovering contexts by clustering the events so that each set of events is executed in the same context. The approach uses literal names and affiliations of people as features and the contents of the documents since these aspects are essential for defining a context. The proposed solution was evaluated on real-life data.

4.1.8 Interactive - Background knowledge

While event logs contain specific information about process executions, they do not contain the background knowledge that is necessary for precisely defining process models. Such background knowledge is often known by analysts and domain experts. Process discovery techniques that are interactive and allow for inputting such information are presented in this subsection.

Yürek, Birant, and BİRANT (2018)'s primary objective is to propose an algorithm that will work on large data sets to produce understandable process models. The study introduced an interactive process miner (IPM) algorithm that creates process models from event logs. The algorithm supports different features for adjusting the process model and provides an interactive environment that helps to validate decisions before applying them in the real world.

Traditional process discovery techniques do not use the knowledge of domain experts while discovering the process models from event logs. Dixit, Verbeek, Buijs, and van der Aalst (2018, October) propose an approach that combines the user's expertise and the information given in event logs. Therefore, the method makes it possible to configure the parameters of the discovery algorithm and make the model discovery process interactive. The case study and the evaluation of the proposed technique showed that it surpasses the traditional model discovery methods.

The discussion about bridging the gap between commercial and academic process mining tools is presented by Leemans, Fahland, and van der Aalst (2014, September). Configuring algorithm parameters, discovering maps, and evaluating is referred to as process exploration. The report identifies process exploration aspects and introduces a process exploration method, namely, the Inductive visual Miner (IvM). IvM demonstrates the synthesis of powerful academic techniques integrated with ProM and a user-friendly package that accompanies commercial tools. The conducted case study shows the feasibility of the solution.

Greco, Guzzo, Lupia, and Pontieri (2015)'s proposal for process model discovery is based on the information given in event logs and background knowledge. The proposed constraint-based framework uses prior knowledge for defining precedence constraints. The paper provides an analysis of the computational complexity of the method. The report includes the results of the experimental activities for validating the applicability of the technique.

The participation of domain experts, who are aware of the processes, can enhance the mined model in the model discovery process. Yahya, Song, Bae, Sul, and Wu (2016) propose the technique for discovering a process model based on event data and user knowledge. Domain knowledge can be applied in the discovery process as a constraint that is used to extract the desired behavior. The method was implemented as a ProM plugin.

Ferreira, Szimanski, and Ralha (2012, September) aim to discover a micro-level model for the agents' behavior and identify the fitness of micro-level models with the macro-level description of the business processes. Event data provides a micro-level sequence of events. The macro-level business process descriptions are provided by the business analysts or are obtained from the documentation. A case study shows that the proposed approach can discover agents' behavior in each activity of a business process for which a macro-level model is identified.

Analyzing learning processes and providing knowledge of the processes online is Okoye, Tawil, Naeem, and Lamine (2016)'s primary target. Data is extracted from learning execution environments for improved model discovery and is transformed into an executable format for mining. The study describes a semantic process mining approach that enriches event data from learning processes. The method can predict individual learning patterns and outcomes through further semantic analysis of the discovered models.

Benevento, Dixit, Sani, Aloini, and van der Aalst (2019, September) discuss the suitability and effectiveness of interactive process mining in healthcare environments. The interactive process discovery method allows domain experts to participate and use domain knowledge in the model discovery process from event logs. The evaluation of the method was done on real-life data from an Italian hospital.

Eck, Sidorova, and Van Der Aalst. (2018, June) consider that very often the main focus of artifact-centric process discovery is on the representation of the artifacts. Therefore the paper's primary focus is to provide a strategy to discover interactions between correlated behaviors in artifacts and analyze their performance. The study provides interactive artifact-oriented process discovery using Composite state machine miner (CSM). The solution has been evaluated using real-life data and has been implemented as a ProM plug-in.

Artifact-centric process models aim to describe complex processes as collecting interacting artifacts (van Eck, Sidorova, and van der Aalst, 2017). In (van Eck, Sidorova, and van der Aalst, 2017) the focus is on how discovered artifacts are related to each other. Moreover, the paper provides visualization and qualification of the interactions between the discovered artifact-centric models. The approach was implemented as a ProM plug-in.

4.1.9 Clustering

Extremely large and complex business process models can be hard to understand. In order to make discovered process models more comprehensible, trace clustering can be applied.

According to (García-Bañuelos, Dumas, La Rosa, De Weerdt, & Ekanayake 2014) trace clustering methods carry drawbacks that are related to process model complexity and understandability. Process discovery methods based on trace clustering produce small and individual models but do not necessarily strive to reduce the size of the set of discovered models. Moreover, those sets of models share duplicated fragments, which makes the process models hard to comprehend. Another drawback of the method is that it produces models with low accuracy. García-Bañuelos, Dumas, La Rosa, De Weerdt, & Ekanayake (2014) proposes a divide-and-conquer process discovery technique that addresses the drawbacks and tries to improve the discovered model concerning fitness and understandability. The method was validated on real-life logs.

Kalenkova, Burattin, Leoni, Aalst, and Sperduti (2018) introduce an approach for discovering multi-perspective hierarchical BPMN models. The paper integrates various process mining techniques. The objective of the work is to fill the gap between consolidated process mining techniques and a wide range of BPMN-compliant tools.

To avoid spaghetti-like process description of the model Folino, Greco, Guzzo, and Pontieri (2011), in their study, focused on applying trace clustering methods to event logs. More precisely, the paper aims to make clustering techniques aware of outlier traces and find predictive models for clustering results. The proposal was validated on real-life application scenarios.

Delias, Doumpos, Grigoroudis, Manolitzas, and Matsatsinis (2015) aims to support decision-making by providing understandable process models in flexible environments, such as healthcare, customer service. The method considered in the study identifies clusters of various traces and connects process discovery methods to subsets of behaviors, and as a result the derived model is clear and understandable. For validation purposes, the technique was applied to a case of a healthcare institution.

Ferreira and Alves (2011, August) focus on discovering communities (social networks) by analyzing the organization's event logs. The paper describes using the hierarchical clustering method with the modularity concept for analyzing communities derived from large event data. The technique defines the division of social networks into collections of clusters. The proposal was implemented as a ProM plugin and a case study was applied.

Cordes, Vogelgesang, and Appelrath (2014, September) target to analyze discovered process models from various viewpoints in multidimensional process mining. Therefore, the clustering method is applied for deriving sets of traces according to the event data attributes. The paper discusses how to spot differences between the almost identical models and introduces a generic approach for calculating and visualizing differences between process models.

Health services research (HSR), a scientific field, analyzes personal health services concerning quality and efficiency. According to Vogelgesang and Appelrath (2013, March) health care processes can be discovered by multidimensional process mining. The paper introduces an approach that improves the understandability of the discovered model by clustering process models.

Song, Günther, and Van der Aalst (2008, September) introduced a divide-and-conquer approach to improve process mining results in flexible environments such as health care services. The approach uses the trace clustering technique, which splits event logs into homogeneous subsets, and for each subgroup, a process model is discovered. The method was integrated with ProM, and for demonstration purposes, a case study was carried out.

Process discovery strives to construct comprehensible and accurate models. De Weerdt, Vanden Broucke, Vanthienen, and Baesens (2013) aim to improve process discovery using trace clustering technique. The study bridges the gap between clustering bias and evaluation bias. The assessment of the proposal was done on real-life logs.

The primary goal of the discussion presented by Ekanayake, Dumas, García-Bañuelos, and La Rosa (2013) is to improve the understanding of the discovered process model. The paper states that the collection of models produced with clustering strategies may be accompanied by excess complexity. The proposed approach splits the models into variants using trace clustering and sub-processes in order to reduce complexity. The technique allows users to put a complexity limit on the model. The approach was tested on real-life logs, and the result showed that the discovered models were reduced in size compared to those discovered by other trace clustering techniques.

De Medeiros, Guzzo, Greco, Van der Aalst, Weijters, Van Dongen, and Sacca (2007, September) focus on improving the discovered model using a trace clustering method. The proposal presents different clustering and process discovery algorithms. The proposed trace clustering algorithm avoids over-generalization, and the discovery algorithm proves to be effective. The solution was implemented in ProM.

(Folino, Guarascio, and Pontieri (2015, June)) discusses process discovery techniques concerning the suitability for BPM systems, such as issue management. The proposal includes a logical event clustering model, a logical trace clustering model, and a collection of workflow schemas. The evaluation was done on real-life data.

According to (Wang, Tang, and Hu 2017, August), clustering techniques may be negligent, which means that sometimes some critical information may be damaged or important activities may be lost. The study introduces a novel approach for clustering event logs. The proposed technique first analyzes the backbone of the event data and then defines constraints used for further trace clustering. For evaluating the applicability of the approach, the experiment on real-life logs was conducted.

4.1.10 Refinement

Automatically discovered process models do not often meet the analysts' expectations because either they are too complex or too standard and do not capture peculiarities of individual processes. Making process discovery interactive so as to allow discovered process refinement is an important research topic.

Yürek, Birant, and BİRANT (2018)'s primary objective is to propose an algorithm that will work on large data sets to produce understandable process models. The study introduced an interactive process miner (IPM) algorithm that creates process models from event logs. The algorithm supports different features for adjusting the process model, including activity deletion, aggregation, and addition operations on the discovered process model. The proposal provides an interactive environment that helps to validate decisions before applying them in the real world.

Traditional process discovery techniques do not use the knowledge of domain experts while discovering the process models from event logs. Dixit, Verbeek, Buijs, and van der Aalst (2018, October) propose an approach that combines the user's expertise and the information given in event logs. Therefore, the method makes it possible to configure the parameters of the discovery algorithm and make the model discovery process interactive. The approach produces sound models and may have duplicate activities. The case study and the evaluation of the proposed technique showed that it surpasses the traditional model discovery methods.

Nguyen, Dumas, Hofstede, La Rosa, and Maggi (2019) discuss an automated stage-based discovery method from event logs to avoid the discovered process model being flat. The method first identifies process stages from event logs and then discovers the model. The alternative of the approach is the divide-and-conquer technique. The divide-and-conquer technique, when applied to real-life logs, produces complex models. The method described in the report outperforms the divide-and-conquer approach concerning model accuracy and complexity.

To avoid discovering spaghetti-like process models Günther, and Van Der Aalst (2007, September) focus on a process mining technique to address this challenge. The proposal is configurable and provides simplified views of a particular process. The study introduces a flexible approach for Fuzzy Mining that adaptively facilitates mined process models.

4.1.11 Software System Logs

While event logs for process discovery usually include logs generated by CRM or ERP systems, process mining techniques can be applied to software runtime logs as well.

Liu (2018) has presented an approach of discovering an understandable hierarchical Petri net model from a software runtime execution log. The proposed approach contributes towards discovering software behavioral models and is proven useful by real-life logs validation. In order to achieve the results, first, the inherently flat execution logs were constructed in a hierarchical manner, and then extended process discovery techniques were utilized to discover software behavior models.

Geng, et al. (2009, September) have worked towards increasing the reach of process mining to the businesses and processes that do not necessarily use computer systems with ready-made event logs. The proposed method consists of two parts - process instance discovery and activity type discovery. The proposed approach is able to map the daily computer operations logs to the conventional event logs that are required for process mining methods.

With the increasing usage of software systems, increased quality is expected from software developers. Özdağoğlu, and Kavuncubaşı (2019) have employed process mining techniques in order to extract learnings about process mapping and performance statistics of a bug-fixing process. The approach has set a goal of revealing bottlenecks that cause risks for the process. The approach was evaluated in a real-life case study and was proven to enable gaining significant knowledge about process steps and roles.

Mazak, and Wimmer (2016) utilize software execution logs and propose execution-based model profiling in order to drive continuous improvement of the designed processes. A prototype of the proposed approach is implemented and validated in a real-life case study of a traffic light system.

Rubin, et al. (2014, September) have applied several process mining methods on different software system logs in the tourism domain. The paper demonstrates that using process mining methods on software enables gaining insights on the real usage of the software and can contribute to usability improvements and software redesign. Moreover, software process models can also be discovered. This work demonstrates that user interactions can be mined from almost every software system.

4.1.12 Configurable

Configurable process models represent a family of processes rather than a single specific process. Discovery of such models became more relevant with the rise of SaaS software usage.

Van Der Aalst (2010, October) has introduced one of the innovative approaches at that time by trying to create configurable and executable process models. This work is important since it allows growing the reach of the process mining discipline by moving process models into “executable reality” from “paper tigers” (Van Der Aalst, 2010, October). This approach enables more organizations to get engaged in BPM and process mining, and also is an enabler for cross-organizational process mining.

Based on the concept of RM_WF_Net, Zeng, et al. (2013) have researched how process mining can be applied for workflow integration use cases. To this end, the authors have defined four coordination patterns and used these patterns to obtain a cross-organizational workflow based on the discovered models of individual organizations.

Assy, van Dongen, and van der Aalst (2017, June) introduce a decomposition-driven approach that allows discovering generic process models from a collection of event logs collected by different process variants. The discovery results are easily understandable and can be browsed at different levels of abstraction. Such configurability of the result view makes it easier for end-users to see commonalities and differences between different process variants. The approach was integrated with ProM and evaluated on both real-life and synthetic logs.

Suriadi, et al. (2014, July) have taken on the complex challenge of conducting a case study to analyze the differences in patient treatments by different hospitals. For this goal, the authors have applied process mining techniques in order to discover different variants of processes conducted in different hospitals while treating patients with similar symptoms. This work presents important insights and guidelines on cross-organizational process mining.

Liu, et al. (2016) have worked towards making cross-organizational process mining more privacy-oriented. The proposed approach allows organizations to discover their private and public process models separately. Then, the public models from across organizations are taken by a neutral middleware party and a new, combined, cross-organizational public process model is created. This model can then be used by companies to combine its desired fragments with their private business process models, hence both preserving privacy and gaining benefit from cross-organizational knowledge.

Buijs, van Dongen, and van der Aalst (2011, August) have also contributed to cross-organizational process mining. Based on the fact that there is increased usage of Software-as-a-Service (SaaS) infrastructures, the authors have introduced an innovative approach of comparing the process models and event logs of different organizations that use the same product (SaaS providers). The approach was validated by comparing real-life processes of different Dutch municipalities.

4.1.13 Multi-perspective

In addition to control flow, process model discovery techniques should be able to discover data attributes associated with activities. Such techniques are referred to as multi-perspective process discovery techniques.

Van Eck, Sidorova, and van der Aalst (2016, September) focus on identifying different facets of models in process discovery. The proposal considers states of various perspectives and studies their relation. Therefore, the approach discovers state-based models highlighting the links of the states. The method was implemented as a ProM plugin.

According to Li, Ge, Huang, Hu, Wu, Yang, ... and Luo (2016), existing process mining algorithms consider only event logs in the model discovery process. The paper introduces a novel concept of token-based log that makes it possible to track the changes of process resources. The study compares the feasibility of the token logs to event logs. For validating the approach, several modeling and mining platform plugins have been developed.

Van der Aalst, Guo, and Gorissen (2015) introduce a notion of process cubes responsible for considering event data from different angles and dimensions. The paper discusses comparative process mining concerning process cubes. The feasibility of the technique was tested on real-life data.

The approach that unites the control-flow and data perspectives is presented by Shraga, Gal, Schumacher, Senderovich, and Weidlich. The combination of the two was done by the extension of the inductive process discovery. In the proposal context, data takes priority over control flow in the model discovery process. In the resulting model, some operations carry data semantics. The method was evaluated on real-life and synthetic logs.

To satisfy particular requirements for the application in financial audits Werner, M., & Gehrke, N. (2015) present a process mining algorithm that improves process audits. The proposed algorithm uses data dependencies for identifying control flow and data flow. The study claims that the presented algorithm as a special purpose algorithm is suitable for financial audits. Still, the approach that combines data and control flow may be used in other application domains. The approach was validated on real-life data.

4.1.14 General

This subsection contains a collection of works that play an important role in process mining but do not fall in any of the aforementioned categories.

Stroiński, Dwornikowski, and Brzeziński (2016) present an innovative distributed algorithm called dRMA that is able to discover process models from Communicating Resource Systems (CRS) event logs. The features of CRS such as its directed communication channels, hierarchy, and resource perspective can be expressed by a Communication Net that was defined by the authors (Stroiński, Dwornikowski, and Brzeziński, 2016). The proposed approach is evaluated on a cluster and the results show significant improvement in the discovery times.

Sun, and Bauer (2016) improve upon existing business process discovery algorithms in order to increase their fitness. By inheriting the basic idea behind the mining algorithm enhancement-based strategy, the authors were able to come up with a process discovery algorithm that is able to discover high-fitness models from real-life event logs.

Roci, and Davidrajuh (2018) present an Alpha-algorithm for process model discovery that is able to discover models with a Petri net representation. The algorithm is special because of its speed, efficiency, and simplicity. The algorithm is implemented in practice and tested on a real case study with a medium-sized event log.

Yano, Nomura, and Kanai (2013, September) have taken on the challenge of making process mining more widely adopted in real-life scenarios. To that end, the work proposes a more practical approach to process mining where firstly a regular transaction database is used instead of event logs. Then, this information is visualized and analyzed, without discovering conventional process models. Lastly, exceptional as well as typical instances of processes are analyzed. The approach was implemented in a tool called BPM-E and was evaluated on two real-life case studies.

Process mining has touched a lot of different domains, including product management. In particular, Bernard, and Andritsos (2017) have used conventional process mining approaches in order to discover and analyze how customers use company products. That is to say, the authors have proposed a way of creating customer journey maps using process model discovery methods (Bernard, and Andritsos, 2017). This work is important because it further widens the reach of the process mining discipline.

He, et al. (2018) propose an α-FL algorithm that is able to discover free-loop structures in process models from incomplete event logs. The resulting discovered process models are represented with Petri nets. The efficiency and correctness of the proposed algorithm are evaluated on real-life case studies and experiments.

Since the complexity of the ILP Miner significantly increases with increased event logs and distinct activities, Verbeek, and van der Aalst (2012, September) have proposed a way of splitting up large event logs using so-called passages. The approach of splitting large event logs and then mining them is evaluated on seven real-life and one synthetic log.

Engel, et al. (2016) propose an approach for mining the Electronic Data Interchange (EDI) messages using conventional process mining techniques. This work paves the way towards discovering inter-organizational business processes and evaluating the performance of such processes. The feasibility of the approach is evaluated on real-life EDI data of a manufacturing company.

Juneja, Kundra, and Sureka (2016, June) have applied process map discovery techniques on the Issue Tracking System (ITS) event logs. This has enabled process discovery of software bug life cycles. In addition, to improve the readability, fitness, and simplicity of discovered process models, the event logs were split into homogeneous subsets. The approach is able to handle bug self-loops, reopening, and back and forth events.

(Augusto, et al. 2016, November) is another paper that contributes to the process model discovery problem but this time the authors have taken an innovative approach rather than the conventional process mining. In order to discover understandable yet accurate process models, Augusto, et al. (2016, November) have reimagined the order of things. First, they discover a model that is as accurate as possible, even though it might not be well structured, and then the resulting accurate model is transformed into a more organized one. real-life evaluation has shown that this order of things outperforms conventional approaches in terms of both accuracy and complexity.

When it comes to the real-life application of process model discovery, it is impossible to avoid the fact that event logs contain exceptional behavior. Traditional discovery algorithms such as ILP-based ones are too accurate and try to perfectly match the event log. To tackle the problem of such overfitting, van Zelst, van Dongen, and van der Aalst (2016, September) propose a filtering technique that is able to discover more understandable and adequate process models which capture only dominant behavior. The technique was integrated with ProM.

In order to increase the reach of process mining and enabling process discovery for more human-centric and ad-hoc processes, Kudo, Ishida, and Sato (2013, December) have introduced a new algorithm that uses a skeleton-extraction procedure and is able to discover better-structured business process models from unstructured workflows. The approach was evaluated and proven effective in the real-life event logs of an insurance company.

Leemans, Fahland, and van der Aalst (2013, June) have contributed to process mining development by presenting an extensible framework that is able to discover sound and fir block-structured process models from any event log. Moreover, the authors have defined the minimum information that is required in the log to rediscover a process model. The proposed polynomial-time algorithm was integrated in ProM.

Van Der Aalst, and Van Dongen (2013) present several approaches for discovering Petri nets from event logs, including the α-algorithm, state-based regions, and language-based regions. More importantly, the work presents some theoretical background and requirements for process model discovery. Van Der Aalst, and Van Dongen (2013) discuss the importance of handling incomplete and noisy logs while discovering process models, thus defining the process of mining open challenges.

Sztyler, et al. (2015) have taken an innovative approach to using the massive data provided by smartphones and smartwatches for health self-tracking. The authors propose applying process discovery techniques on the self-tracking data in order to provide visibility and insights to individuals on what their daily processes look like. The paper presents interesting challenges and conclusions for future research.

Werner (2017) has worked on process model discovery in the financial context. The scope of finances and accounting has allowed using data dependencies of accounting structures as the source of discovering the order of events in the control flow. As opposed to the conventional, temporal orders and timestamp comparisons for discovering the order of events, the data dependency approach presented by Werner (2017) has proven to be less complex based on three different real-life case studies.

Heuristics Miner is one of the well-known process discovery algorithms. It also has well-known disadvantages affecting its reliability. vanden Broucke, and De Weerdt (2017) have extended upon the Heuristics Miner with a strong focus on its robustness and flexibility and presented a new algorithm that has better performance, can handle duplicate tasks, and allows more flexible configuration options (vanden Broucke, and De Weerdt, 2017). The presented algorithm is known by the name “Fodina”.

With the goal of streamlining clinical pathways in the healthcare sector, De Weerdt, et al. (2012, May) has proposed a methodology for intelligent analysis of clinical pathway data using process mining approaches. Regardless of the complexity involved, the paper proposes the methodology that enables extracting tangible insights from healthcare data by using both drill up and drill down perspectives.

Alizadeh, and Norani (2018) have developed a new algorithm for process model discovery called ICMA. In addition to the actual process discovery phase, the algorithm also includes preprocessing and postprocessing stages for event logs. These additional stages have enabled the ICMA algorithm to handle noisy logs well enough. The algorithm was evaluated on a multitude of logs and compared to conventional algorithms such as the Vazquise algorithm. The results show that the ICMA algorithm has superior precision and completeness over its predecessors.

van Zelst, van Dongen, and van der Aalst (2015, June) have built on top of the existing Integer Linear Programming (ILP) formulation in order to develop a new process discovery algorithm that unifies two types of language-based regions. Moreover, van Zelst, van Dongen, and van der Aalst (2015, June) have presented a generalized ILP objective function that helps to find more suitable processes during discovery.

Sánchez-Charles, et al. (2016), propose a scalable technique that solves the problem of underfitting in process discovery algorithms. More specifically, according to Sánchez-Charles, et al. (2016), the same activity occurring a different number of times in the traces act differently, and they have come up with a way to detect such peculiarities. The proposed approach was validated and compared with existing benchmarks at the time.

De Cnudde, Claes, and Poels (2014) improve upon the well-known Heuristics Miner by updating the tool as well as the underlying algorithm. The improved technique was integrated with ProM as a new artifact and was evaluated on several different logs. The evaluation results show that the validity and completeness of the updated miner are superior.

Leemans, Tax, and ter Hofstede (2018, October) have combined the Inductive Miner, the Evolutionary Tree Miner, the Local Process Model Miner, and the new bottom-up recursive technique with the goal of an improved overall algorithm for process model discovery. Evaluation results show that the combination of algorithms amplifies their advantages, while the disadvantages of individual algorithms are in some cases balanced out by other ones. Evaluation on real-life logs has shown that in some cases, the new, combined algorithm surpassed its competition.

With the increasing scale of globally interconnected businesses and the use of software in everyday operations, the amount of data that is the input for process mining has significantly grown. In addition, in most cases, different systems in the organization are disconnected and independent. Redlich, et al. (2014, May) have proposed an approach for scalable dynamic process discovery in order to tackle these challenges. The proposed solution enables keeping up with the scaling and dynamics of businesses.

Yang, et al. (2017, August) introduce a new alignment-guided state-splitting hidden Markov models inference algorithm (AGSS) for process model discovery. According to the evaluation results, the proposed algorithm is more accurate, efficient and discovers more readable process models than its predecessors.

Mokhov, Carmona, and Beaumont (2016) show how Conditional Partial Order Graphs (CPOGs) can be used in process mining, more specifically in discovering readable process models. The representation with the help of CPOGs can be used to reveal hidden links between data and control flow in the process. The evaluation shows that the proposed approach can be used for effective and efficient visualization of discovery results.

Dustdar, Hoffmann, and Van der Aalst (2005) tackle the problem of mining ad-hoc processes. This work is important since ad-hoc processes are closer to the real-life situations of how businesses operate. The paper demonstrates the use of their approach using Caramba, a process-aware collaboration system that allows ad-hoc processes. Dustdar, Hoffmann, and Van der Aalst (2005) use EMiT, MinSoN and Teamlog in order to achieve their goals.

Wen, et al. (2007) focus on mining non-free-choice constructs during process discovery. Moreover, the paper shows that there are two kinds - explicit and implicit dependencies between tasks. The proposed algorithm is integrated with ProM and its evaluation results show significant improvements over existing algorithms at the time.

Weber, et al. (2015, April) improve upon existing process mining techniques for process discovery and conformance checking in order to handle sub-processes that are instantiated multiple times during process execution. The proposed approach was evaluated on real-life event logs.

Carmona, and Cortadella (2013) propose using numerical abstract domains in process discovery in order to tackle the problems of large logs and the formal properties of the discovered model. The proposed technique enables discovering general process models that do not require knowledge. The approach was integrated with ProM.

Song, et al. (2015) propose an innovative process discovery technique that leverages activity dependencies in traces and is able to discover concurrencies even in incomplete logs. the proposed approach was evaluated on both real-life and synthetic logs and has proven its significance.

Baker, et al. (2017) have used process mining and more specifically process discovery techniques in order to discover the actual process of the clinical pathways for cancer patients going through chemotherapy. The paper aimed at establishing a reproducible process mining method, use its outputs to quantify clinical pathways for patients, and use these insights to do a cost-effectiveness analysis of the current process. The results of the study were proven to be beneficial.

Leemans, van der Aalst, and van den Brand (2018, March) present an innovative hierarchy and recursion extension to the process tree model along with the recursion aware discovery algorithm. The proposed technique allows for efficiently discovering and analyzing the operational characteristics of software systems. The approach was integrated with ProM and validated on real-life logs. The results show that the approach is feasible and useful.

Leemans, Fahland, and van der Aalst (2015) have discovered a framework along with three algorithms that are both scalable and discover sound models. The proposed framework computes a directly-follows graph and applies the divide-and-conquer strategy in order to handle large logs. The experiments show that the algorithms can cope with event logs of 100 000 000 traces and 10 000 activities.

Song, et al. (2017) have introduced a method for scientific process (workflow) mining that supports both intra-cloud and inter-cloud scientific workflow mining. The proposed approach was integrated with ProM and evaluated on real-life processes. The experiments show that the method is both effective and efficient.

Due to the increased usage of Software-as-a-Service and Cloud Computing by the organizations, the processes and the logs that the organizations follow can often be similar. Buijs, van Dongen, and van der Aalst (2011, August) have utilized this fact and introduced an innovative approach of comparing the process models and event logs of different organizations that use the same infrastructure. The approach was validated by comparing processes of different Dutch municipalities.

Anand, Gupta, and Sureka (2015, December) have improved upon the Fuzzy Miner (FM) algorithm and made it Utility-Based to efficiently mine process models driven by utility thresholds. The proposed approach allows end-users to define how utility is measured (in terms of profit, value, quantity…). The approach is validated on a real-life dataset and its effectiveness is demonstrated.

The following subsections will present process discovery related works, clustered based on common aspects they touch upon.

4.1.14.1 Genetic

This subsection presents a collection of works that tackle the problem of process model discovery by utilizing genetic algorithms.

Buijs, van Dongen, and van der Aalst (2012, June) are the first ones to propose an innovative process discovery algorithm that allows search processes to be guided by a user's preferences while still remaining correct. The innovative algorithm falls under the umbrella of genetic process mining algorithms and tackles the problems that its predecessors had (producing anomalous models, problems with simplicity, generalization, precision…). The proposed algorithm allows balancing different quality dimensions while ensuring the soundness of the model.

Yang, et al. (2015 June) have filled the research gap by creating a general algorithm that is capable of handling diverse logs. The proposed algorithm first divides the whole event log into several sub-logs. Then these sub-logs are separately mined by conventional process discovery algorithms. Finally, the several resulting process models are assembled together with the help of a genetic algorithm-based optimizer while balancing the four quality dimensions (Yang, et al., 2015 June). The approach was evaluated on both synthetic and real-life logs and produced successful results.

Ghazal, Ghoniemy, and Salama (2019) propose an innovative way of looking at process model discovery as a Multi-Objective Optimization Problem. As such, they employ the NSGA-II (one of the most widely used Multi-Objective Optimizers) algorithm for process model discovery. Evaluation on real-life logs showed that the proposed method outperforms its predecessors.

Buijs, Van Dongen, and van Der Aalst (2014) present the Evolutionary Tree Miner (ETM) algorithm that allows end-users to put weights on the four quality dimensions and steer the discovery process according to their preferences. Moreover, the paper shows that all four quality dimensions (replay fitness, precision, generalization, and simplicity) are important for process discovery, but considering only the latter three makes sense if replay fitness is satisfactory.

Zhao, Liu, and Dai (2014) introduce a genetic programming approach to process mining which simplifies the resulting discovered process models compared to the existing algorithms. The authors also propose a new metric for measuring process complexity, which takes into consideration the role of cohesion and coupling. Higher fitness from this new metric also provides guidelines for process model redesign.

Van Der Aalst, Buijs, and Van Dongen (2011, June) have tackled the challenge of internally inconsistent process model discovery by proposing a tree representation that will ensure model soundness. By implementing a genetic algorithm that discovers process trees, the authors were able to reduce the search space, thus resulting in more sound process models and efficient process discovery.

4.1.14.2 Preprocessing

Properly processed and prepared event logs are a prerequisite for precise and understandable process model discovery. This subsection will present works that introduce different preprocessing techniques involved in business process discovery.

Zhu, et al. (2019) have introduced a software process activity classifier that enables building event-activity mappings from software development event streams. The approach is used to extract SVN logs and map each of the events with business activities. The technique was applied to two real-life software development process logs and the results show that the software process activities can be mined automatically and in real-time from SVN logs.

Baier, Mendling, and Weske (2014) propose an approach of preprocessing event logs such that they conform to the same abstraction level that is needed by the business. Domain knowledge extracted from process documentation was used to map process events with business activities. The resulting approach also supports concurrency and was evaluated in two real-life case studies.

Bala, et al. (2018, October) fill in the research gap by proposing an event log preprocessing method that is able to infer case and activity identifiers. This work contributes to making process mining more applicable to real-life scenarios since it is often the case that the event logs do not contain case or event identifiers that are crucial for process discovery algorithms. The results of this work (Bala, et al., 2018, October) are evaluated in an industry case study.

Similar work was done by Pourmirza, Dijkman, and Grefen (2017). The authors have introduced the correlation miner - a technique that is able to first, preprocess event logs by determining which events are associated with the same case identifiers, and then run a conventional process discovery algorithm. The proposed technique was evaluated on both synthetic and real-life logs and was proven feasible.

Soares, Santoro, and Baião (2013) describe a detailed method for discovering relevant information for Knowledge-Intensive processes based on informal communication channels such as emails. The extracted emails are preprocessed in order to extract relevant information and then the text is mined for relevant data. The proposed approach was validated in a real-life case study.

Wang, et al. (2012) propose an innovative λ-algorithm that aims at improving the efficiency of process model discovery. The approach is able to deliver satisfactory results by eliminating some of the conventional steps of process mining and treating event logs in a completely different way. First, the events and their occurrence frequency are extracted from the log, ignoring their order. Secondly, the order relations are extracted, making it possible to do the final step of discovering process models with ordering relations.

Geng, et al. (2009, September) have worked on a method to preprocess artifacts of daily computer operations as standard event logs. This work is important as it allows for increasing the reach of process mining to the businesses and processes that do not necessarily use computer systems with ready-made event logs. The proposed method consists of two parts - process instance discovery and activity type discovery.

Repta, and Stanescu (2017, May) propose a system architecture in order to tackle the needs for properly representing event data, associating events with process instances, and determining higher-level actions that correspond to the events. From a high-level perspective, the proposed method is a way for preprocessing the logs.

In order to effectively preprocess event logs or determine if an event log is suitable for the Heuristic Miner, Kurniati, Kusuma, and Wisudiawan (2016) have defined a set of characteristics for the event logs that are best suited for this process discovery algorithm. The authors conducted a separate research for each case study and reported the analysis of heuristic miner results for different types of event logs.

Sani, van Zelst, and van der Aalst (2017, September) introduce a practical method for effectively preprocessing event logs with the aim of removing excessive noise from them. The described general purpose filtering method was integrated with both ProM and RapidProM and was evaluated on both synthetic and real-life logs. The results show that the technique indeed accurately removes noise from event logs and hence improves process discovery.

To tackle the problem of anonymizing business processes, Irshad, et al. (2015, July) have developed a privacy-preserving Business Process Recommendation and Composition System (BPRCS). This system processes event logs in order to anonymize them. Then, the anonymized logs are mined in order to end up with similarly masked business processes that can be securely shared between organizations. The proposed approach was validated in real-life experiments.

Syamsiyah, van Dongen, and van der Aalst (2017, September) have reduced the time that analysts have to wait for exporting and preprocessing event logs before running process mining algorithms on them. This was made possible by creating a method that preprocesses events on the fly, as the events are recorded in the system, as opposed to preprocessing them in a batch before process mining. The preprocessed events are persisted in a database and can be accessed by the analyst whenever desired, without large waiting times. The approach was integrated with ProM and validated in real-life use cases.

4.1.14.3 Incomplete Logs

A lot of real-life logs are incomplete. In order for process discovery to be realistic and applicable to real-world processes, the algorithms need to be able to handle incomplete event logs.

Leemans, Fahland, and van der Aalst (2014, June) have studied the effects of incomplete logs on process discovery. The authors first evaluate the impact of log incompleteness on behavioral relations and then introduce probabilistic behavioral relations that are less sensitive to incompleteness. Using these new relations, a more robust process discovery algorithm was introduced, which was able to rediscover a process of the original system from incomplete logs.

4.1.14.4 Low-Level Abstraction

Abstracting away from low-level, detailed activities is essential to make discovered process models understandable for stakeholders. The works related to such techniques are presented in this subsection.

Mannhardt, et al. (2018) propose a method for Guided Process Discovery (GPD) which helps discover high-level process models using low-level event logs. The low-level events are first grouped based on similar behavioral patterns. These groups are then "translated" to a higher level, abstracted activities that are understandable for stakeholders. The resulting high-level log is then checked for conformance against the initial low-level event log. The proposed approach was validated on two real-life event logs and the results show that process models discovered this way are more comprehensible and relevant for answering business-oriented questions.

Ferreira, Szimanski, and Ralha (2014) have also tried to fill in the gap between low-level steps recorded in the event logs and high-level activities of a business process. By introducing an approach to mine mappings between the low-level runtime events with high-level process activities, the authors made it possible to generate suggestions on how the process model can be extended to capture the behavior described in the event log. The approach was evaluated on a real-life event log and the results show that it can improve the model step-by-step until all the behavior recorded in the event log is covered.

Tax, et al. (2016, September) also contribute to the work of abstracting low-level logs of smart home environments to a higher level, more comprehensible activities. This abstraction helps the resulting discovered models to be more understandable and precise. The paper also presents a framework that can automatically abstract low-level sensor events to the higher, human activity level by utilizing machine learning. The proposed approach was validated on real-life smart home event logs and the results show that process discovery after low-level abstraction is indeed more comprehensible and precise.

Greco, Guzzo, and Pontieri (2008) have improved upon classical process discovery methods by introducing a method for abstracting process models. This approach makes it possible to capture discovered models on different granularity (abstraction) levels based on the end-use case. The approach was integrated with ProM and validated on real-life cases, ending up with promising results.

4.1.14.5 Large logs

The works that tackle the problem of large logs are presented in this subsection. The proposed techniques improve the process mining performance while discovering models from excessive logs.

Verbeek, van der Aalst, and Munoz-Gama (2017) have tackled the problem of handling large logs for process model discovery. In their proposed approach, a big discovery problem is divided into smaller, more manageable problems and is called "Divide and Conquer". The technique was integrated with ProM and the evaluation results show that it provides significant speed benefits over other existing methods. It is important to note that the paper also discusses how decomposition might lead to different discovered models, but that this can even turn out to be a positive side effect.

Hompes, Verbeek, and van der Aalst (2014, November) have set a goal of defining quality measures for decomposed process mining methods. The authors have defined three quality notions that can assess decomposition before actual discovery or conformance checking algorithms are run. Moreover, using these defined quality measures, the paper provides an algorithm that is able to find high-quality decomposition in little time (Hompes, Verbeek, and van der Aalst, 2014, November). As a result, this paper effectively contributes to the improvement of decomposition process mining.

4.1.14.6 Invisible tasks

Real-life event logs often implicitly contain activities that are carried out during process execution. Since these activities are not explicitly present, they are referred to as “invisible”. Being able to discover such activities is an important challenge in process mining.

Guo, et al. (2016, September) have tackled the problem of invisible tasks involved in non-free-choice constructs in process mining by the algorithm called α$. The algorithm introduces a new ordering relation between tasks and solves the aforementioned problem in this manner. The algorithm was integrated with ProM and evaluated on real-life logs. The results have shown that α$ is a significant improvement on process mining algorithms.

Wen, et al. (2010) have also worked on improving the process discovery algorithms to handle invisible tasks. In particular, the authors introduce an algorithm α#, which is able to detect invisible tasks of type Initialize, Skip, Redo, Switch or Finalize. In addition, a new ordering relation is introduced that detects false dependencies between the tasks. The α# algorithm was evaluated on both synthetic and real-life logs and showed promising results.

4.1.14.7 Non-atomic

For the discovered process models to accurately capture underlying processes, discovery algorithms should distinguish concurrency and interleaving. Discovery techniques that allow such distinction are presented in this subsection. These techniques handle event logs that contain non-atomic activities. i.e., for each activity, its start and completion time is known.

Wen, et al. (2009) propose an innovative approach for process mining based on the “Start” and “Complete” event types. This approach helps discover parallelism in process execution. In addition, the proposed algorithm overcomes the limitations of its predecessors, thus contributing to the overall improvement of process mining.

Leemans, Fahland, and van der Aalst (2016, September) propose an algorithm for process model discovery, the performance of which can be measured more accurately. Moreover, the method proposed in this paper can handle life cycle data and is able to distinguish concurrency and interleaving (Leemans, Fahland, and van der Aalst, 2016, September).

4.1.14.8 Incremental

Business processes are rarely finite. They are usually continuous and the event logs retrieved at any given time do not fully describe the process. Hence, it is important to be able to incrementally update discovered process models once new execution logs become available.

Sun, et al. (2006, October) present a process discovery algorithm that is able to handle process models with optional tasks and also is able to incrementally update the discovered process model once the new process execution logs become available. The proposed algorithm can also simplify the comparison of two process models (Sun, et al., 2006, October). The approach is evaluated on real-life logs.

Ferilli, De Carolis, and Redavid (2013, June) also present an algorithm for incremental process mining and show how it can be applied in the everyday work of organizations and employees. The incremental process discovery algorithm proposed by Ferilli, De Carolis, and Redavid (2013, June) is a good technique for predicting daily routines and needs for the users, as well as comparing the predicted processes with the actual ones. The approach has been evaluated on both synthetic and real-life cases,

Similar work on incremental process mining was done by Ferilli, Redavid, and Esposito (2015, September). The authors present a First-Order Logic incremental method for process discovery. The effectiveness and efficiency of the algorithm are evaluated in both synthetic and real-life datasets.

Ferilli (2013) describes the workflow management tool WoMan and the entire algorithmic apparatus behind it. The tool is able to incrementally mine business processes and quickly learn new behavior even in case of noise and changed behavior, among other benefits. It also provides the ability to export/import process models from/into standard representations (Petri nets), simulate, and monitor process models. WoMan process discovery algorithm also has an improved expressive power compared to its predecessors. The tool was evaluated on both synthetic and real-life logs.

4.1.14.9 Unstructured

This subsection introduces the work related to process model discovery from unstructured logs.

Banziger, Basukoski, and Chaussalet (2018, June) propose an original framework for discovering process models from unstructured CRM data. This is made possible by using the Latent Dirichlet Allocation (LDA), which automatically detects and assigns labels to activities (Banziger, Basukoski, and Chaussalet, 2018, June). This work is important because it increases the reach for traditional process mining methods by making it possible to handle unstructured data such as those provided by CRM systems. The proposed approach was evaluated on real-life CRM data and the results prove the feasibility of the method.

4.1.14.10 Probabilistic

Attempts of utilizing a probabilistic approach such as Bayesian belief networks to process model discovery are presented in this subsection.

Vasilecas, Savickas, and Lebedys (2014, October) present an approach of extracting directed acyclic graphs from event logs. Experiments show that the proposed approach is feasible for real-life scenarios (Vasilecas, Savickas, and Lebedys, 2014, October). This work is important since it contributes to increasing the usage of Bayesian belief networks in process mining.

Kumar, Bhattacharyya, Varshneya (2010, July) have introduced a process model discovery approach that is able to discover process models from unstructured logs. The approach employs Gaussian mixture models and hidden Markov models. The technique is applied on a set of real-life processes and the work demonstrates that its results are comparable to manually labeling the logs and then mining them using the state-of-the-art discovery algorithm.

4.1.14.11 Email Driven

Often, in real-life scenarios, structured and ready-made event logs are not available. Innovative approaches for process model discovery from email threads are presented in this subsection.

Stuit, and Wortmann (2012) present an innovative approach of process mining by trying to mine email-driven business processes. Their proposed approach identifies email threads from interactions and constructs an interaction-centric process model. Email header fields are used as process-related information. The proposed method is implemented in a tool called Email

Interaction Miner and is evaluated in a real-life business case (Stuit, and Wortmann, 2012). The results show that mining business processes from email threads can bring improvement opportunities and insights to organizations.

4.1.14.12 Simulation Model

In order to make discovered process models more understandable, researchers have worked towards simulating them. Such works are presented in this subsection.

Wang, et al. (2017, December) present a mature tool that applies Data to the Fuzzy-DEVS model (D2FS) method in order to discover discrete event simulation models. The tool is integrated with ProM along with the SimStudio simulation tool and it can simulate the Fuzzy-DEVS model. The tool is evaluated on two real-life case studies.

Wang, et al. (2018) propose a new system inference method. This method allows knowledge extraction from data to represent complex systems. The method consists of first extracting the event logs from data, then discovering a transition system using process discovery techniques, and finally integrating fuzzy methods to automatically generate the Fuzzy-DEVS model. The approach is integrated with ProM and comes with the SimStudio simulator.

4.1.14.13 Noise

Real-life event logs are rarely free of impurities. They often contain noise that usually is not detected by overfitting discovery algorithms. Works that introduce techniques of noise detection and reduction in process discovery are presented in this subsection.

Măruşter, et al. (2006) propose a process model discovery approach that utilizes machine learning in order to predict the relationships between activities and induce rule sets. In addition, the work discusses the effect of noise and imbalance of execution priorities on the process of discovery. Finally, the approach is validated in a real-life case study.

Huang, and Kumar (2012) have also taken on the challenge to improve process discovery algorithms in order to be able to handle noisy (not 100% accurate) event logs. In the paper (Huang, and Kumar, 2012) they share their important findings, including the claim that it is possible to represent any process log using only self-loop and optional structures. In addition, the authors claim that for any given log, there is not a unique but rather multiple process models that satisfies the log. In addition, the paper concludes that a fully accurate process model that contains every trace from the log might have lower quality in the case of noisy logs. Hence, the use of self-loop and optional structures should be controlled in order to balance the model's accuracy and quality. The proposed approach is validated on several synthetic and real-life, noisy logs.

Li, et al. (2018) solves the problem of effectively discovering process models from noisy logs by introducing an innovative approach to detecting noise. The resulting discovered model is represented in a Petri net representation, In order to detect noise, the distance between the traces is calculated. The traces of the smaller clusters are treated as noise, while the larger clusters are mined as usual. The proposed approach showed superior correctness when compared with the α++ algorithm.

Redlich, et al. (2014, September) propose an algorithm for process model discovery that consists of BP-domain constructs as opposed to traditional Petri or Causal nets that can be hard to understand for business analysts. The proposed algorithm is able to handle noise and inaccuracy in logs by having a top-down approach and letting the different constructs compete with each other until the most suitable one is left.

Leemans, Fahland, and van der Aalst (2013, August) present a technique that is able to tackle the problems that its predecessors have. According to the authors, the presented technique is able to cope with rare behaviors (noise) and large logs while ensuring the soundness of the results. The approach was integrated with ProM and compared with existing approaches in terms of quality and performance.

Folino, et al. (2009, September) improve the existing process discovery algorithms for real-world applications. In this attempt, the paper proposes an algorithm that is able to cope with duplicate and hidden tasks, as well as noisy logs and non-free choice relationships among activities. The approach is scalable for real-world cases.

Tax, Sidorova, and van der Aalst (2017) also tackle the problem of discovering process models from noisy and chaotic logs. The authors propose a novel approach of filtering out noisy activities as opposed to standard, frequency-based filtering. The proposed approach has been evaluated on 17 different real-life logs and the results show that the new activity filtering method provides better quality resulting models.

Pecchia, et al. (2020) propose an approach for detecting software failures from application logs using process mining. Since software application logs are inherently noisy, the proposed method consequently contributes to process model discovery from noisy logs. Firstly, process models are discovered from logs and then a conformance checking is run to detect deviations. The evaluation shows mixed results depending on different applications.

Redlich, et al. (2014, November) have implemented a set of modifications for an existing Constructs Competition Miner (CCM). CCM is designed to handle noise and not supported behavior, hence these characteristics are inherited by the proposed algorithm as well. The main goal of this work is to enable dynamic process discovery of a runtime process model.

Weijters, and Ribeiro (2011, April) have created a new process representation language based on Causal nets along with the relevant process mining algorithm. The proposed framework is better able to handle noisy logs. The resulting process models are easier to understand even in the case of non-trivial constructs. The algorithm was integrated with ProM and evaluated on multiple real-life noisy logs.

4.2 Declarative Process Model Discovery

When business processes are large and contain a lot of decision points, procedural process models can become very complex and spaghetti-like. To tackle this problem, declarative process models were introduced. Such models contain a set of business rules that the process must follow instead of specific control-flow. Works related to discovery of declarative process models are introduced in this section.

4.2.1 Sub-processes

Similar to procedural process models, declarative ones can also contain sub-processes for improved understandability. Work related to discovering declarative sub-processes is introduced in this subsection.

Jalali (2014) targets improving existing process model discovery methods in order to be able to discover common aspects from different event logs. The author uses concepts related to Aspect-Oriented Business Process Management (AOBPM) in order to achieve this goal. The paper uses real-life event logs from a bank in order to explain the problem with existing process mining algorithms and then test the new, aspect-oriented method that was developed.

4.2.2 Discovered Process Model Refinement

Pruning and refining discovered process models is an important feature for discovery algorithms. Works contributing to this feature are presented in this subsection.

Maggi, Bose, and van der Aalst (2013 June) have worked towards decluttering discovered business models with their work. The authors (Maggi, Bose, van der Aalst, 2013 June) have introduced various pruning techniques in order to refine declarative business process models that have been discovered from event logs. In addition, the paper (Maggi, Bose, van der Aalst, 2013 June) illustrates how domain knowledge can effectively be used in guiding the process of discovery. The techniques presented by Maggi, Bose, and van der Aalst (2013 June) have been integrated into the process mining tool ProM and have been evaluated on real-life event logs from a Dutch hospital.

In addition to pruning regular (not data-aware) discovered process models, Bose, Maggi, and van der Aalst (2013) have introduced techniques of pruning business constraints from discovered declarative business process models using correlations. Using data attribute correlations and disambiguations allows finding discriminatory patterns, identifying outliers, and analyzing bottlenecks (Bose, Maggi, van der Aalst, 2013). The presented technique has been integrated with ProM and tested with real-life event logs from a Dutch academic hospital.

4.2.3 Multi-perspective

Multi-perspective process discovery in the scope of declarative process models has the same meaning as for the procedural ones. Declarative process discovery techniques that take into consideration the existing data attributes are presented in this subsection.

Sturm, Schönig, and Jablonski (2018) aim at introducing an efficient process mining framework that enables multi-perspective constraint discovery. For this purpose, authors employ a distributed processing method MapReduce, and test their implementation on several real-life event logs.

Leno, et al. (2020) have worked on discovering declarative process models with data conditions. More specifically, the paper (Leno, et al. 2020) aims at discovering constraints with two or more activities, both of which are associated with specific data conditions when the activities occur. Hence, Leno, et al. (2020) tackle discovering multi-perspective process models with correlation conditions. The authors present and compare two approaches for achieving this goal. Testing on the combination of synthetic and real-life logs shows that the first, clustering and rule mining technique outperforms the second, redescription mining technique in terms of re-discovering artificially injected constraints in a log. Redescription mining also has its own advantage of discovering rules with higher confidence and lower support.

Maggi, et al. (2013) have also worked on declaring declarative business process models with data constraints. The technique proposed by (Maggi, et al., 2013) is able to discover data activation conditions during process model discovery. The method has been validated on a real-life event log of a cancer treatment process. One of the major benefits of the method proposed by Maggi, et al. (2013) is that it is able to capture recurrent rules that relate to pairs of activities, making it possible to provide a summary view of key rules that govern the process.

Bose, Maggi, and van der Aalst (2013) employ multi-perspective, data correlation conditions in order to prune discovered declarative process models. The introduced method was validated on real-life event logs and was proved to be efficient in limiting the number of constraints that are shown to only the essential ones, making the discovered process models much easier to understand with data conditions.

Sturm, Fichtner, and Schönig (2019) aim at improving upon existing process mining methods in order to increase their performance. The authors have employed the latest big data mining technology and the distributed processing method MapReduce in order to achieve this goal (Sturm, Fichtner, and Schönig, 2019). The algorithms presented in their work cover the full set of MP-Declare constraints as well as allow a simpler introduction for custom constraints. The method was validated on real-life event logs and showed that the introduced method is indeed able to discover process models in a reasonable amount of time.

Sutrisnowati, et al. (2017) have targeted the problem of scalability and real-life usability of Multi-perspective process model discovery methods. The authors have introduced a scalable indexing algorithm for MP process analysis (Sutrisnowati, et al., 2017). The proposed solution only indexes the attributes in the selected events and focuses on time data constraints. The approach was implemented as a ProM plugin and validated on real-life event logs.

Sturm, Schönig, and Jablonski (2018) propose an efficient technique for mining declarative process models with data attributes. By building on top of the distributed processing method MapReduce, the authors were able to develop an innovative multi-perspective discovery algorithm that is more efficient than its predecessors (Sturm, Schönig, and Jablonski, 2018). The proposed approach has been tested on real-life event logs.

4.2.4 Clustering

Works presented in this subsection contribute to improving the understandability of discovered declarative process models by utilizing clustering approaches.

Bernardi, Cimitile, and Mercaldo (2018) propose a cloud architecture that helps improve resource management distribution for organizations. The authors have discovered an approach of continuously extracting the data from the software logs and applying online process mining techniques to analyze these logs (Bernardi, Cimitile, and Mercaldo 2018). The resulting process models are declarative and provide insight on different variants of executed business processes. The approach was validated in a real-life process of online shopping.

Bernardi, Cimitile, and Maggi (2014) introduce a cloud computing multi-tenancy architecture that supports cross-organizational process executions. The authors have also developed an approach for the systematic extraction of data from system logs in order to organize them in an ordinary event log fashion (Bernardi, Cimitile, and Maggi, 2014). Finally, the authors have integrated online process mining techniques in order to extract business rules from software process logs (Bernardi, Cimitile, and Maggi, 2014). The resulting method is able to discover different variants of the process execution to provide greater insights to organizations.

4.2.5 Online

Being able to handle a continuous stream of events and discover declarative process models in real-time is an important research area. Works related to this topic are presented in this subsection.

Bernardi, Cimitile, and Mercaldo (2018) have discovered a systematic approach to extracting and serializing software execution runtime logs into event logs. The resulting event logs are streamed to online process mining tools in order to discover multiple variants of the process execution (Bernardi, Cimitile, and Mercaldo 2018). This paper is one of the best examples of how online process mining techniques can be used and the benefits they can bring to organizations.

Similarly, Bernardi, Cimitile, and Maggi (2014) have introduced a cloud architecture that allows for the effective usage of process runtime logs as event logs. The continuous process execution logs are serialized and used for online process mining, making it possible to extract business rules from process executions in real-time.

Burattin, Cimitile, and Maggi (2014 September) propose a tool for visualization of online process model discovery. Since the process discovery is online, the discovered model is changing in real-time, hence the visualization is referred to as a “process movie” (Burattin, Cimitile, and Maggi, 2014 September). The tool's effectiveness is validated on a real-life log of health insurance claims handling processes in a travel agency.

Burattin, et al. (2015) propose an innovative method of online discovery of declarative process models. With the introduced approach it is possible to store all events over a longer period of time. In addition, the proposed method can provide meaningful information on changes occurring in a process during its execution. The method was validated on both synthetic and real-life event logs.

4.2.6 General

Important works that do not fall under any of the aforementioned subsections are presented in this subsection.

Di Ciccio, and Mecella (2013, April) have developed a two-step algorithm for discovering declarative process models - MINERful++. The proposed algorithm first builds a knowledge base based on the information from execution logs. Then, the knowledge base is queried and the statistical support of constraints is computed (Di Ciccio, and Mecella, 2013, April). MINERful++ algorithm is important because it is efficient, independent, and capable of removing unnecessary, subsumed constraints.

Schönig, et al. (2016, June) utilize relational database RelationalXES and introduce an innovative mining approach with SQL. The benefit of the method described in this paper is that the mining is customizable and not restricted to the predefined control-flow constraints. The presented approach was validated on several real-life event logs.

Maggi, Bose, and van der Aalst (2012, June) have developed an Apriori algorithm for discovering declarative business process models. The algorithm reduces the search space, making the discovery more fast and efficient. In addition, the discovered models are pruned and the results are highly readable. The proposed approach was tested on real-life event logs.

Räim, et al. (2014 October) have developed a proof of concept for temporal logic query checking. This concept stands in between standard process discovery and conformance checking. Declare constraints are discovered from actual event logs and then compared to the constraints pre-defined by a user. In addition to providing a true-false type answer for process compliance, the method is able to provide a detailed view on which traces violate or satisfy the constraints (Räim, et al., 2014 October). The approach was validated on both real-life and synthetic logs.

Di Ciccio, and Mecella (2015) present an approach to discovering declarative process models for artful processes. For this purpose, the authors have developed a two-phase algorithm called MINERful which operates in a similar way as described in (Di Ciccio, and Mecella, 2013, April). In addition to actual algorithm development, the paper provides a thorough theoretical and empirical evaluation of the approach in terms of its performance and quality based on a real case study. The method is validated on both synthetic and real-life event logs.

Maggi, Mooij, and Van der Aalst (2011, April) present one of the early works of declarative process mining. The paper presents an approach to automatically discover Declare process models that were integrated with the mining tool ProM. The proposed approach was validated on real-life logs from the maritime safety and security domain.

The following subsections will present process discovery related works, grouped based on common aspects.

4.2.6.1 Disjunction

Since declarative process models describe processes using a set of business rules, being able to handle different sorts of complex, including target-branched declare constraints is an important feature for discovery algorithms.

Di Ciccio, Maggi, and Mendling (2014) tackle the problem of complex declarative models' readability with their work. More specifically, they define the class of Target-Branched Declare constraints and examine its properties. In addition to the theoretical analysis, the authors provide a technique for efficiently discovering Target-Branched Declare models. The proposed technique is validated on both synthetic and real-life logs.

Di Ciccio, Maggi, and Mendling (2016) improve upon their previous (2014) work in an attempt to discover branched Declare constraints efficiently. The paper includes theoretical discussions as well as a technique implemented in practice that was validated on both synthetic and real-life logs (Di Ciccio, Maggi, and Mendling, 2016). In addition to the actual usability of the method developed by Di Ciccio, Maggi, and Mendling (2016), their work provides interesting insights and theory that helps in understanding Target-Branched Declare.

4.2.6.2 Efficiency Algorithm

While discovery algorithms might be able to discover precise and understandable process models, their performance is also an important subject. Works that contribute to improving the performance of declarative process discovery algorithms are presented in this subsection.

Kala, et al. (2016 September) use a combination of the Apriori algorithm and a group of algorithms for Sequence Analysis with a goal of improving the performance of an existing Declare Miner. Declare Miner is a plugin for ProM that is able to discover declarative process models from event logs. The validation on both synthetic and real-life logs has shown that Kala, et al. (2016 September) were able to significantly improve the performance of the existing Declare Miner plugin.

Maggi, et al. (2018) improves on the previous works of the authors and aims to further improve the performance of the Declare Miner. The proposed approach was developed in a way that allows parallelization using two different partitioning methods - search space partitioning and database partitioning. The new, improved Declare Miner was integrated with the newer version of the ProM tool and the validation on both synthetic and real-life logs show that the performance has indeed improved.

4.2.6.3 Email Driven

Similar to section 4.1.14.11, this section presents works that contribute to discovering declarative process models from email threads.

Di Ciccio, and Mecella (2012, May) aim to develop a mining algorithm that is able to efficiently compute the constraints that characterize artful processes. The processes in the form of workflows are obtained from the tool called MailOfMine, which aims to automatically build workflows based on the flow of email messages (Di Ciccio, and Mecella, 2012, May). The resulting algorithm is validated on real-life workflows.

4.2.6.4 Multi Instance

Being able to see similarities and differences between different instances of the same process is important for analysts. Among other benefits, it allows them to clearly see violations of predefined constraints.

Winter, Stertz, and Rinderle-Ma (2020) aim at discovering Instance Spanning Constraints (ISCs) from event logs. Since ISCs are the instruments that allow establishing controls across multiple instances of the process, discovering them could be very beneficial to businesses. The paper compares four different discovery algorithms for ISC discovery from process execution logs (Winter, Stertz, and Rinderle-Ma, 2020). The discovered ISCs are put in context for domain experts to validate and refine them further. The results are validated on both synthetic and real-life logs.

4.3 Hybrid Process Model Discovery

This section provides an overview of how different works have answered the business-oriented question “How are the instances of a business process executed?” Different aspects related to this question are discussed in the following subsections.

4.3.1 Sub-processes

Similar to procedural and declarative process model discovery, it is important to be able to discover subprocesses for hybrid models as well.

Maggi, Slaats, and Reijers (2014, September) propose an approach for discovering hybrid process models from event logs. The resulting process model consists of multiple sub-processes that are either declarative or procedural. The approach is integrated with ProM and validated on real-life event logs from a financial context (Maggi, Slaats, and Reijers, 2014, September). This work is innovative and important since it allows leveraging the benefits of both procedural and declarative process models, resulting in a more precise and understandable final model.

4.3.2 General

This subsection contains a collection of works that do not fall in any of the aforementioned categories.

Schunselaar, et al. (2018 July) present a technique for discovering hybrid (procedural + declarative) process models from event logs. First, the authors use Inductive Miner for discovering a fully procedural (imperative) log, and then they “balance out” its imprecise parts with declarative models. As a result, this approach increases the precision of discovered process models without overfitting the data. The technique was integrated with ProM and validated on both synthetic and real-life logs.

5 Conclusion

In the scope of this thesis paper, a Systematic Literature Review on the topic of process mining was conducted. The SLR protocol was defined and a value-driven framework for process mining use cases was described based on the SLR results. Moreover, SLR results were presented in more detail by summarizing the process mining-related papers that have contributed to business process discovery, one of the process mining methods belonging to the value of transparency.

A total of 839 papers were analyzed and grouped into different value categories. 267 papers fell under transparency, out of which 183 papers about process model discovery were summarized in this work.

For the future, other process mining use cases under the transparency value such as process model enhancement, social network mining, goal modeling, and decision mining need to be considered as well as papers that contribute to other values such as efficiency, compliance, quality, and agility.

6 References

[1] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1, p. 2). Heidelberg: Springer.

[2] Palmer N. (n.d.). What is BPM? Retrieved June 02, 2021 from https://bpm.com/what-is-bpm

[3] Franz, P., & Kirchmer, M. (2013). Value-Driven Business Process Management: The Value-Switch for Lasting Competitive Advantage (1st ed.). McGraw-Hill Education.

[4] Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004), 1-26.

[5] Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, xiii-xxiii.

[6] Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F. M., Marrella, A., ... & Soo, A. (2018). Automated discovery of process models from event logs: Review and benchmark. IEEE transactions on knowledge and data engineering, 31(4), 686-705.

[7] Dunzer, S., Stierle, M., Matzner, M., & Baier, S. (2019, June). Conformance checking: a state-of-the-art literature review. In Proceedings of the 11th international conference on subject-oriented business process management (pp. 1-10).

[8] Di Francescomarino, C., Ghidini, C., Maggi, F. M., & Milani, F. (2018, September). Predictive process monitoring methods: Which one suits me best?. In International Conference on Business Process Management (pp. 462-479). Springer, Cham.

[9] Kitchenham, B. A., Mendes, E., & Travassos, G. H. (2007). Cross versus within-company cost estimation studies: A systematic review. IEEE Transactions on Software Engineering, 33(5), 316-329.

[10] Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software, 80(4), 571-583.

[11] Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. Informing Science, 9.

[12] Fink, A. (2019). Conducting research literature reviews: From the internet to paper. Sage publications.

[13] Kofod-Petersen, A. (2012). How to do a structured literature review in computer science. Ver. 0.1. October, 1.

[14] Badakhshan, P., Conboy, K., Grisold, T., & vom Brocke, J. (2019). Agile business process management: A systematic literature review and an integrated framework. Business Process Management Journal.

[15] Naderifar, V., Sahran, S., & Shukur, Z. (2019). A review on conformance checking technique for the evaluation of process mining algorithms. TEM Journal, 8(4), 1232.

[16] Nuritha, I., & Mahendrawathi, E. R. (2019). Behavioural Similarity Measurement of Business Process Model to Compare Process Discovery Algorithms Performance in Dealing with Noisy Event Log. Procedia Computer Science, 161, 984-993.

[17] Okoli, C. (2015). A guide to conducting a standalone systematic literature review. Communications of the Association for Information Systems, 37(1), 43.

[18] van Eck, M. L., Sidorova, N., & van der Aalst, W. M. (2017, July). Guided interaction exploration in artifact-centric process models. In 2017 ieee 19th conference on business informatics (cbi) (Vol. 1, pp. 109-118). IEEE.

[19] X. Lu, M. Nagelkerke, D. v. d. Wiel and D. Fahland, (2015, November) Discovering Interacting Artifacts from ERP Systems," in IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 861-873, 1 Nov.-Dec. 2015, doi: 10.1109/TSC.2015.2474358.

[20] Popova, V., Fahland, D., & Dumas, M. (2015, March). Artifact Lifecycle Discovery. International Journal of Cooperative Information Systems, 24(01), 1550001. doi:10.1142/s021884301550001x

[21] Popova, V., & Dumas, M. (2014). Discovering Unbounded Synchronization Conditions in Artifact-Centric Process Models. Business Process Management Workshops Lecture Notes in Business Information Processing, 28-40. doi:10.1007/978-3-319-06257-0_3

[22] Eck, M. L., Sidorova, N., & Wil M. P. Van Der Aalst. (2018, June). Guided Interaction Exploration and Performance Analysis in Artifact-Centric Process Models. Business & Information Systems Engineering, 61(6), 649-663. doi:10.1007/s12599-018-0546-0

[23] Augusto, A., Conforti, R., Dumas, M., La Rosa, M., & Bruno, G. (2018). Automated discovery of structured process models from event logs: The discover-and-structure approach. Data & Knowledge Engineering, 117, 373-392. doi:10.1016/j.datak.2018.04.007

[24] Kalenkova, A., Burattin, A., Leoni, M. D., Aalst, W. V., & Sperduti, A. (2018). Discovering high-level BPMN process models from event data. Business Process Management Journal, 25(5), 995-1019. doi:10.1108/bpmj-02-2018-0051

[25] Conforti, R., Dumas, M., García-Bañuelos, L., & La Rosa, M. (2016). BPMN Miner: Automated discovery of BPMN process models with hierarchical structure. Information Systems, 56, 284-303. doi:10.1016/j.is.2015.07.004

[26] Augusto, A., Conforti, R., Dumas, M., La Rosa, M., & Polyvyanyy, A. (2018). Split miner: Automated discovery of accurate and simple business process models from event logs. Knowledge and Information Systems, 59(2), 251-284. doi:10.1007/s10115-018-1214-x

[27] Conforti, R., Augusto, A., La Rosa, M., Dumas-Menijvar, M., & Garcia-Banuelos, L. (2016). BPMN Miner 2.0: Discovering hierarchical and block-structured BPMN process models. In Proceedings of the BPM Demo Track 2016 [CEUR Workshop Proceedings, Volume 1789 (pp. 39-43). Sun SITE Central Europe.

[28] Wang, Y., Wen, L., Yan, Z., Sun, B., & Wang, J. (2015). Discovering BPMN Models with Sub-processes and Multi-instance Markers. Lecture Notes in Computer Science On the Move to Meaningful Internet Systems: OTM 2015 Conferences, 185-201. doi:10.1007/978-3-319-26148-5_11

[29] Al-Ali, H., Cuzzocrea, A., Damiani, E., Mizouni, R., & Tello, G. (2019). A composite machine-learning-based framework for supporting low-level event logs to high-level business process model activities mappings enhanced by flexible BPMN model translation. Soft Computing, 24(10), 7557-7578. doi:10.1007/s00500-019-04385-6

[30] Li, J., Bose, R. P., & Wil M. P. Van Der Aalst. (2011). Mining Context-Dependent and Interactive Business Process Maps Using Execution Patterns. Business Process Management Workshops Lecture Notes in Business Information Processing, 109-121. doi:10.1007/978-3-642-20511-8_10

[31] Hassani, M., Siccha, S., Richter, F., & Seidl, T. (2015). Efficient Process Discovery From Event Streams Using Sequential Pattern Mining. 2015 IEEE Symposium Series on Computational Intelligence. doi:10.1109/ssci.2015.195

[32] Ferreira, D. R., & Thom, L. H. (2012). A semantic approach to the discovery of workflow activity patterns in event logs. International Journal of Business Process Integration and Management, 6(1), 4. doi:10.1504/ijbpim.2012.047909

[33] Leemans, M., & Wil M. P. Van Der Aalst. (2015). Discovery of Frequent Episodes in Event Logs. Lecture Notes in Business Information Processing Data-Driven Process Discovery and Analysis, 1-31. doi:10.1007/978-3-319-27243-6_1

[34] Deeva, G., & Weerdt, J. D. (2019). Understanding Automated Feedback in Learning Processes by Mining Local Patterns. Business Process Management Workshops Lecture Notes in Business Information Processing, 56-68. doi:10.1007/978-3-030-11641-5_5

[35] Leno, V., Armas-Cervantes, A., Dumas, M., La Rosa, M., & Maggi, F. M. (2018). Discovering process maps from event streams. Proceedings of the 2018 International Conference on Software and System Process. doi:10.1145/3202710.3203154

[36] Redlich, D., Molka, T., Gilani, W., Blair, G. S., & Rashid, A. (2014, November). Scalable Dynamic Business Process Discovery with the Constructs Competition Miner. In SIMPDA (pp. 91-107).

[37] Burattin, A., Sperduti, A., & Aalst, van der, W. M. P. (2012). Heuristics miners for streaming event data. (arXiv.or; Vol. 1212.6383 [cs.DB]). s.n.

[38] Burattin, A., Sperduti, A., & Wil M. P. Van Der Aalst. (2014). Control-flow discovery from event streams. 2014 IEEE Congress on Evolutionary Computation (CEC). doi:10.1109/cec.2014.6900341

[39] Syamsiyah, A., Dongen, B. F., & Wil M. P. Van Der Aalst. (2018). Recurrent Process Mining with Live Event Data. Business Process Management Workshops Lecture Notes in Business Information Processing, 178-190. doi:10.1007/978-3-319-74030-0_13

[40] Zelst, S. J., Dongen, B. F., & Wil M. P. Van Der Aalst. (2017). Event stream-based process discovery using abstract representations. Knowledge and Information Systems, 54(2), 407-435. doi:10.1007/s10115-017-1060-2

[41] Nguyen, H. H., La Rosa, M., Dumas-Menijvar, M., & ter Hofstede, A. (2017). Stage-based business process mining. In Proceedings of the Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017)(CEUR Workshop Proceedings, Volume 1848) (pp. 161-169). Sun SITE Central Europe.

[42] Leemans, M., & Wil M. P. Van Der Aalst. (2017). Modeling and Discovering Cancelation Behavior. On the Move to Meaningful Internet Systems. OTM 2017 Conferences Lecture Notes in Computer Science, 93-113. doi:10.1007/978-3-319-69462-7_8

[43] Kalenkova, A. A., & Lomazova, I. A. (2014). Discovery of Cancellation Regions within Process Mining Techniques. Fundamenta Informaticae, 133(2-3), 197-209. doi:10.3233/fi-2014-1071

[44] Nguyen, H., Dumas, M., Hofstede, A. H., La Rosa, M., & Maggi, F. M. (2019). Stage-based discovery of business process models from event logs. Information Systems, 84, 214-237. doi:10.1016/j.is.2019.05.002

[45] Dongen, B. F., & Adriansyah, A. (2010). Process Mining: Fuzzy Clustering and Performance Visualization. Business Process Management Workshops Lecture Notes in Business Information Processing, 158-169. doi:10.1007/978-3-642-12186-9_15

[46] Nguyen, H., Dumas, M., ter Hofstede, A. H., La Rosa, M., & Maggi, F. M. (2017, June). Mining business process stages from event logs. In International Conference on Advanced Information Systems Engineering (pp. 577-594). Springer, Cham.

[47] Molka, T., Redlich, D., Drobek, M., Zeng, X., & Gilani, W. (2015). Diversity Guided Evolutionary Mining of Hierarchical Process Models. Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. doi:10.1145/2739480.2754765

[48] Conforti, R., Dumas, M., García-Bañuelos, L., & La Rosa, M. (2014). Beyond Tasks and Gateways: Discovering BPMN Models with Subprocesses, Boundary Events and Activity Markers. Lecture Notes in Computer Science Business Process Management, 101-117. doi:10.1007/978-3-319-10172-9_7

[49] Ekanayake, C. C., Dumas, M., García-Bañuelos, L., & La Rosa, M. (2013). Slice, Mine and Dice: Complexity-Aware Automated Discovery of Business Process Models. Lecture Notes in Computer Science Business Process Management, 49-64. doi:10.1007/978-3-642-40176-3_6

[50] Sun, Y., & Bauer, B. (2016). A Graph and Trace Clustering-based Approach for Abstracting Mined Business Process Models. Proceedings of the 18th International Conference on Enterprise Information Systems. doi:10.5220/0005833900630074

[51] Weber, I., Farshchi, M., Mendling, J., & Schneider, J. G. (2015, April). Mining processes with multi-instantiation. In Proceedings of the 30th Annual ACM Symposium on Applied Computing (pp. 1231-1237).

[52] Jalali, A. (2014, September). Aspect mining in business process management. In International Conference on Business Informatics Research (pp. 246-260). Springer, Cham.

[53] Yzquierdo-Herrera, R., Silverio-Castro, R., & Lazo-Cortés, M. (2013). Sub-process discovery: Opportunities for process diagnostics. In Enterprise Information Systems of the Future (pp. 48-57). Springer, Berlin, Heidelberg.

[54] Štajner, T., Mladenić, D., & Grobelnik, M. (2010). Exploring contexts and actions in knowledge processes. In Proceedings of the 2nd Workshop on Context, Information and Ontologies.

[55] Yürek, I., Birant, D., & BİRANT, K. U. (2018). Interactive process miner: a new approach for process mining. Turkish Journal of Electrical Engineering & Computer Sciences, 26(3), 1314-1328.

[56] Dixit, P. M., Verbeek, H. M. W., Buijs, J. C., & van der Aalst, W. M. (2018, October). Interactive data-driven process model construction. In International conference on conceptual modeling (pp. 251-265). Springer, Cham.

[57] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2014, September). Exploring processes and deviations. In International Conference on Business Process Management (pp. 304-316). Springer, Cham.

[58] Greco, G., Guzzo, A., Lupia, F., & Pontieri, L. (2015). Process discovery under precedence constraints. ACM Transactions on Knowledge Discovery from Data (TKDD), 9(4), 1-39.

[59] Yahya, B. N., Song, M., Bae, H., Sul, S. O., & Wu, J. Z. (2016). Domain-driven actionable process model discovery. Computers & Industrial Engineering, 99, 382-400.

[60] Ferreira, D. R., Szimanski, F., & Ralha, C. G. (2012, September). A hierarchical Markov model to understand the behaviour of agents in business processes. In International Conference on Business Process Management (pp. 150-161). Springer, Berlin, Heidelberg.

[61] Okoye, K., Tawil, A. R. H., Naeem, U., & Lamine, E. (2016). Discovery and enhancement of learning model analysis through semantic process mining. International Journal of Computer Information Systems and Industrial Management Applications, 8(2016), 93-114.

[62] Benevento, E., Dixit, P. M., Sani, M. F., Aloini, D., & van der Aalst, W. M. (2019, September). Evaluating the effectiveness of interactive process discovery in healthcare: a case study. In International Conference on Business Process Management (pp. 508-519). Springer, Cham.

[63] Folino, F., Greco, G., Guzzo, A., & Pontieri, L. (2011). Mining usage scenarios in business processes: Outlier-aware discovery and run-time prediction. Data & Knowledge Engineering, 70(12), 1005-1029.

[64] García-Bañuelos, Dumas, La Rosa, De Weerdt, & Ekanayake (2014). Controlled automated discovery of collections of business process models. Information Systems, 46, 85-101.

[65] Delias, P., Doumpos, M., Grigoroudis, E., Manolitzas, P., & Matsatsinis, N. (2015). Supporting healthcare management decisions via robust clustering of event logs. Knowledge-Based Systems, 84, 203-213.

[66] Ferreira, D. R., & Alves, C. (2011, August). Discovering user communities in large event logs. In International Conference on Business Process Management (pp. 123-134). Springer, Berlin, Heidelberg.

[67] Cordes, C., Vogelgesang, T., & Appelrath, H. J. (2014, September). A generic approach for calculating and visualizing differences between process models in multidimensional process mining. In International Conference on Business Process Management (pp. 383-394). Springer, Cham.

[68] Vogelgesang, T., & Appelrath, H. J. (2013, March). Multidimensional process mining: a flexible analysis approach for health services research. In Proceedings of the Joint EDBT/ICDT 2013 Workshops (pp. 17-22).

[69] Song, M., Günther, C. W., & Van der Aalst, W. M. (2008, September). Trace clustering in process mining. In International conference on business process management (pp. 109-120). Springer, Berlin, Heidelberg.

[70] De Weerdt, J., Vanden Broucke, S., Vanthienen, J., & Baesens, B. (2013). Active trace clustering for improved process discovery. IEEE Transactions on Knowledge and Data Engineering, 25(12), 2708-2720.

[71] De Medeiros, A. K. A., Guzzo, A., Greco, G., Van der Aalst, W. M., Weijters, A. J. M. M., Van Dongen, B. F., & Sacca, D. (2007, September). Process mining based on clustering: A quest for precision. In International conference on business process management (pp. 17-29). Springer, Berlin, Heidelberg.

[72] Wang, P., Tang, A., & Hu, K. (2017, August). A novel trace clustering technique based on constrained trace alignment. In International Conference on Human Centered Computing (pp. 53-63). Springer, Cham.

[73] Folino, F., Guarascio, M., & Pontieri, L. (2015, June). Mining multi-variant process models from low-level logs. In International Conference on Business Information Systems (pp. 165-177). Springer, Cham.

[74] Günther, C. W., & Van Der Aalst, W. M. (2007, September). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In International conference on business process management (pp. 328-343). Springer, Berlin, Heidelberg.

[75] van Eck, M. L., Sidorova, N., & van der Aalst, W. M. (2016, September). Discovering and exploring state-based models for multi-perspective processes. In International Conference on Business Process Management (pp. 142-157). Springer, Cham.

[76] Li, C., Ge, J., Huang, L., Hu, H., Wu, B., Yang, H., ... & Luo, B. (2016). Process mining with token carried data. Information Sciences, 328, 558-576.

[77] van der Aalst, W. M. P., Guo, S., & Gorissen, P. J. B. (2015). Comparative process mining in education: an approach based on process cubes.

[78] Shraga, R., Gal, A., Schumacher, D., Senderovich, A., & Weidlich, M. (2019, June). Inductive Context-aware Process Discovery. In 2019 International Conference on Process Mining (ICPM) (pp. 33-40). IEEE.

[79] Werner, M., & Gehrke, N. (2015). Multilevel process mining for financial audits. IEEE Transactions on Services Computing, 8(6), 820-832.

[80] Liu, C. (2018). Automatic discovery of behavioral models from software execution data. IEEE Transactions on Automation Science and Engineering, 15(4), 1897-1908.

[81] Geng, L., Buffett, S., Hamilton, B., Wang, X., Korba, L., Liu, H., & Wang, Y. (2009, September). Discovering structured event logs from unstructured audit trails for workflow mining. In International Symposium on Methodologies for Intelligent Systems (pp. 442-452). Springer, Berlin, Heidelberg.

[82] Özdağoğlu, G., & Kavuncubaşı, E. (2019). Monitoring the software bug‐fixing process through the process mining approach. Journal of Software: Evolution and Process, 31(7), e2162.

[83] Mazak, A., & Wimmer, M. (2016). On Marrying Model-driven Engineering and Process Mining: A Case Study in Execution-based Model Profiling. In SIMPDA (pp. 78-88).

[84] Rubin, V. A., Mitsyuk, A. A., Lomazova, I. A., & van der Aalst, W. M. (2014, September). Process mining can be applied to software too!. In Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (pp. 1-8).

[85] Van Der Aalst, W. M. (2010, October). Configurable services in the cloud: Supporting variability while enabling cross-organizational process mining. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 8-25). Springer, Berlin, Heidelberg.

[86] Zeng, Q., Sun, S. X., Duan, H., Liu, C., & Wang, H. (2013). Cross-organizational collaborative workflow mining from a multi-source log. Decision support systems, 54(3), 1280-1301.

[87] Assy, N., van Dongen, B. F., & van der Aalst, W. M. (2017, June). Discovering hierarchical consolidated models from process families. In International Conference on Advanced Information Systems Engineering (pp. 314-329). Springer, Cham.

[88] Suriadi, S., Mans, R. S., Wynn, M. T., Partington, A., & Karnon, J. (2014, July). Measuring patient flow variations: A cross-organisational process mining approach. In Asia-Pacific Conference on Business Process Management (pp. 43-58). Springer, Cham.

[89] Liu, C., Duan, H., Zeng, Q., Zhou, M., Lu, F., & Cheng, J. (2016). Towards comprehensive support for privacy preservation cross-organization business process mining. IEEE Transactions on Services Computing, 12(4), 639-653.

[90] Stroiński, A., Dwornikowski, D., & Brzeziński, J. (2016). A distributed discovery of communicating resource systems models. IEEE Transactions on Services Computing, 12(2), 172-185.

[91] Sun, Y., & Bauer, B. (2016). A Novel Fitness Improvement Method for Mined Business Process Models. In CAiSE Forum (pp. 25-32).

[92] Roci, A., & Davidrajuh, R. (2018). A polynomial-time alpha-algorithm for process mining. International Journal of Simulation: Systems, Science and Technology, 19(5), 12-1.

[93] Yano, K., Nomura, Y., & Kanai, T. (2013, September). A practical approach to automated business process discovery. In 2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops (pp. 53-62). IEEE.

[94] Bernard, G., & Andritsos, P. (2017). A process mining based model for customer journey mapping. In Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017) (Vol. 1848, pp. 49-56). CEUR Workshop Proceedings.

[95] He, Z., Du, Y., Wang, L. U., Qi, L., & Sun, H. (2018). An alpha-FL algorithm for discovering free loop structures from incomplete event logs. IEEE Access, 6, 27885-27901.

[96] Verbeek, H. E., & van der Aalst, W. M. (2012, September). An experimental evaluation of passage-based process discovery. In International Conference on Business Process Management (pp. 205-210). Springer, Berlin, Heidelberg.

[97] Engel, R., Krathu, W., Zapletal, M., Pichler, C., Bose, R. J. C., van der Aalst, W., ... & Huemer, C. (2016). Analyzing inter-organizational business processes. Information Systems and e-Business Management, 14(3), 577-612.

[98] Juneja, P., Kundra, D., & Sureka, A. (2016, June). Anvaya: An algorithm and case-study on improving the goodness of software process models generated by mining event-log data in issue tracking systems. In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC) (Vol. 1, pp. 53-62). IEEE.

[99] Augusto, A., Conforti, R., Dumas, M., La Rosa, M., & Bruno, G. (2016, November). Automated discovery of structured process models: Discover structured vs. discover and structure. In International Conference on Conceptual Modeling (pp. 313-329). Springer, Cham.

[100] van Zelst, S. J., van Dongen, B. F., & van der Aalst, W. M. (2016, September). Avoiding over-fitting in ILP-based process discovery. In International conference on business process management (pp. 163-171). Springer, Cham.

[101] Kudo, M., Ishida, A., & Sato, N. (2013, December). Business process discovery by using process skeletonization. In 2013 International Conference on Signal-Image Technology & Internet-Based Systems (pp. 976-982). IEEE.

[102] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2013, June). Discovering block-structured process models from event logs-a constructive approach. In International conference on applications and theory of Petri nets and concurrency (pp. 311-329). Springer, Berlin, Heidelberg.

[103] Van Der Aalst, W. M., & Van Dongen, B. F. (2013). Discovering petri nets from event logs. In Transactions on Petri nets and other models of concurrency vii (pp. 372-422). Springer, Berlin, Heidelberg.

[104] Sztyler, T., Völker, J., Carmona Vargas, J., Meier, O., & Stuckenschmidt, H. (2015). Discovery of personal processes from labeled sensor data: An application of process mining to personalized health care. In Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data: Brussels, Belgium, June 22-23, 2015 (pp. 31-46). CEUR-WS. org.

[105] Werner, M. (2017). Financial process mining-Accounting data structure dependent control flow inference. International Journal of Accounting Information Systems, 25, 57-80.

[106] vanden Broucke, S. K., & De Weerdt, J. (2017). Fodina: a robust and flexible heuristic process discovery technique. decision support systems, 100, 109-118.

[107] De Weerdt, J., Caron, F., Vanthienen, J., & Baesens, B. (2012, May). Getting a grasp on clinical pathway data: an approach based on process mining. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 22-35). Springer, Berlin, Heidelberg.

[108] Alizadeh, S., & Norani, A. (2018). ICMA: A new efficient algorithm for process model discovery. Applied Intelligence, 48(11), 4497-4514.

[109] van Zelst, S. J., van Dongen, B. F., & van der Aalst, W. M. (2015, June). ILP-Based Process Discovery Using Hybrid Regions. In ATAED@ petri nets/ACSD (pp. 47-61).

[110] Sánchez-Charles, D., Solé, M., Carmona, J., & Muntés-Mulero, V. (2016). Improving Process Model Precision by Loop Unrolling. In SIMPDA (pp. 89-99).

[111] De Cnudde, S., Claes, J., & Poels, G. (2014). Improving the quality of the Heuristics Miner in ProM 6.2. Expert Systems with Applications, 41(17), 7678-7690.

[112] Leemans, S. J., Tax, N., & ter Hofstede, A. H. (2018, October). Indulpet miner: Combining discovery algorithms. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 97-115). Springer, Cham.

[113] Redlich, D., Gilani, W., Molka, T., Drobek, M., Rashid, A., & Blair, G. (2014, May). Introducing a framework for scalable dynamic process discovery. In Enterprise Engineering Working Conference (pp. 151-166). Springer, Cham.

[114] Yang, S., Zhou, M., Chen, S., Dong, X., Ahmed, O., Burd, R. S., & Marsic, I. (2017, August). Medical workflow modeling using alignment-guided state-splitting HMM. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 144-153). IEEE.

[115] Mokhov, A., Carmona, J., & Beaumont, J. (2016). Mining conditional partial order graphs from event logs. In Transactions on Petri Nets and Other Models of Concurrency XI (pp. 114-136). Springer, Berlin, Heidelberg.

[116] Dustdar, S., Hoffmann, T., & Van der Aalst, W. (2005). Mining of ad-hoc business processes with TeamLog. Data & Knowledge Engineering, 55(2), 129-158.

[117] Wen, L., Van Der Aalst, W. M., Wang, J., & Sun, J. (2007). Mining process models with non-free-choice constructs. Data Mining and Knowledge Discovery, 15(2), 145-180.

[118] Weber, I., Farshchi, M., Mendling, J., & Schneider, J. G. (2015, April). Mining processes with multi-instantiation. In Proceedings of the 30th Annual ACM Symposium on Applied Computing (pp. 1231-1237).

[119] Carmona, J., & Cortadella, J. (2013). Process discovery algorithms using numerical abstract domains. IEEE Transactions on Knowledge and Data Engineering, 26(12), 3064-3076.

[120] Song, W., Jacobsen, H. A., Ye, C., & Ma, X. (2015). Process discovery from dependence-complete event logs. IEEE Transactions on Services Computing, 9(5), 714-727.

[121] Baker, K., Dunwoodie, E., Jones, R. G., Newsham, A., Johnson, O., Price, C. P., ... & Hall, G. (2017). Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. International journal of medical informatics, 103, 32-41.

[122] Leemans, M., van der Aalst, W. M., & van den Brand, M. G. (2018, March). Recursion aware modeling and discovery for hierarchical software event log analysis. In 2018 IEEE 25th international conference on software analysis, evolution and reengineering (SANER) (pp. 185-196). IEEE.

[123] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2015). Scalable process discovery with guarantees. In Enterprise, Business-Process and Information Systems Modeling (pp. 85-101). Springer, Cham.

[124] Song, W., Chen, F., Jacobsen, H. A., Xia, X., Ye, C., & Ma, X. (2017). Scientific workflow mining in clouds. IEEE Transactions on Parallel and Distributed Systems, 28(10), 2979-2992.

[125] Buijs, J. C., van Dongen, B. F., & van der Aalst, W. M. (2011, August). Towards cross-organizational process mining in collections of process models and their executions. In International Conference on Business Process Management (pp. 2-13). Springer, Berlin, Heidelberg.

[126] Anand, K., Gupta, N., & Sureka, A. (2015, December). Utility-based control flow discovery from business process event logs. In International Conference on Big Data Analytics (pp. 69-83). Springer, Cham.

[127] Stuit, M., & Wortmann, H. (2012). Discovery and analysis of e-mail-driven business processes. Information Systems, 37(2), 142-168.

[128] Buijs, J. C., van Dongen, B. F., & van der Aalst, W. M. (2012, June). A genetic algorithm for discovering process trees. In 2012 IEEE Congress on Evolutionary Computation (pp. 1-8). IEEE.

[129] Yang, L., Kang, G., Cai, W., & Zhou, Q. (2015, June). An effective process mining approach against diverse logs based on case classification. In 2015 IEEE International Congress on Big Data (pp. 351-358). IEEE.

[130] Ghazal, M. A., Ghoniemy, S., & Salama, M. A. (2019). Multi-Objective Optimization for Automated Business Process Discovery. In KDIR (pp. 89-104).

[131] Buijs, J. C., van Dongen, B. F., & van der Aalst, W. M. (2014). Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. International Journal of Cooperative Information Systems, 23(01), 1440001.

[132] Zhao, W., Liu, X., & Dai, W. (2014). Simplified process model discovery based on role-oriented genetic mining. The Scientific World Journal, 2014.

[133] Van Der Aalst, W., Buijs, J., & Van Dongen, B. (2011, June). Towards improving the representational bias of process mining. In International Symposium on Data-Driven Process Discovery and Analysis (pp. 39-54). Springer, Berlin, Heidelberg.

[134] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2014, June). Discovering block-structured process models from incomplete event logs. In International Conference on Applications and Theory of Petri Nets and Concurrency (pp. 91-110). Springer, Cham.

[135] Sun, W., Li, T., Peng, W., & Sun, T. (2006, October). Incremental Workflow Mining with Optional Patterns. In 2006 IEEE International Conference on Systems, Man and Cybernetics (Vol. 4, pp. 2764-2771). IEEE.

[136] Ferilli, S., De Carolis, B., & Redavid, D. (2013, June). Logic-based incremental process mining in smart environments. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 392-401). Springer, Berlin, Heidelberg.

[137] Ferilli, S., Redavid, D., & Esposito, F. (2015, September). Logic-based incremental process mining. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 218-221). Springer, Cham.

[138] Ferilli, S. (2013). Woman: Logic-based workflow learning and management. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(6), 744-756.

[139] Guo, Q., Wen, L., Wang, J., Yan, Z., & Philip, S. Y. (2016, September). Mining invisible tasks in non-free-choice constructs. In International Conference on Business Process Management (pp. 109-125). Springer, Cham.

[140] Wen, L., Wang, J., van der Aalst, W. M., Huang, B., & Sun, J. (2010). Mining process models with prime invisible tasks. Data & Knowledge Engineering, 69(10), 999-1021.

[141] Verbeek, H. M. W., van der Aalst, W. M., & Munoz-Gama, J. (2017). Divide and conquer: A tool framework for supporting decomposed discovery in process mining. The Computer Journal, 60(11), 1649-1674.

[142] Hompes, B. F. A., Verbeek, H. M. W., & van der Aalst, W. M. (2014, November). Finding suitable activity clusters for decomposed process discovery. In International Symposium on Data-Driven Process Discovery and Analysis (pp. 32-57). Springer, Cham.

[143] Mannhardt, F., de Leoni, M., Reijers, H. A., van der Aalst, W. M., & Toussaint, P. J. (2018). Guided process discovery–a pattern-based approach. Information systems, 76, 1-18.

[144] Ferreira, D. R., Szimanski, F., & Ralha, C. G. (2014). Improving process models by mining mappings of low-level events to high-level activities. Journal of Intelligent Information Systems, 43(2), 379-407.

[145] Tax, N., Sidorova, N., Haakma, R., & van der Aalst, W. (2016, September). Mining process model descriptions of daily life through event abstraction. In Proceedings of SAI intelligent systems conference (pp. 83-104). Springer, Cham.

[146] Greco, G., Guzzo, A., & Pontieri, L. (2008). Mining taxonomies of process models. Data & Knowledge Engineering, 67(1), 74-102.

[147] Măruşter, L., Weijters, A. T., Van Der Aalst, W. M., & Van Den Bosch, A. (2006). A rule-based approach for process discovery: Dealing with noise and imbalance in process logs. Data mining and knowledge discovery, 13(1), 67-87.

[148] Huang, Z., & Kumar, A. (2012). A study of quality and accuracy trade-offs in process mining. INFORMS Journal on Computing, 24(2), 311-327.

[149] Li, W., Zhu, H., Liu, W., Chen, D., Jiang, J., & Jin, Q. (2018). An anti-noise process mining algorithm based on minimum spanning tree clustering. IEEE Access, 6, 48756-48764.

[150] Redlich, D., Molka, T., Gilani, W., Blair, G., & Rashid, A. (2014, September). Constructs competition miner: Process control-flow discovery of bp-domain constructs. In International Conference on Business Process Management (pp. 134-150). Springer, Cham.

[151] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2013, August). Discovering block-structured process models from event logs containing infrequent behaviour. In International conference on business process management (pp. 66-78). Springer, Cham.

[152] Folino, F., Greco, G., Guzzo, A., & Pontieri, L. (2009, September). Discovering expressive process models from noised log data. In Proceedings of the 2009 international database engineering & applications symposium (pp. 162-172).

[153] Tax, N., Sidorova, N., & van der Aalst, W. M. (2017). Discovering more precise process models from event logs by filtering out chaotic activities. arXiv preprint arXiv:1711.01287.

[154] Pecchia, A., Weber, I., Cinque, M., & Ma, Y. (2020). Discovering process models for the analysis of application failures under uncertainty of event logs. Knowledge-Based Systems, 189, 105054.

[155] Redlich, D., Molka, T., Gilani, W., Blair, G., & Rashid, A. (2014, November). Dynamic constructs competition miner-occurrence-vs. time-based ageing. In International Symposium on Data-Driven Process Discovery and Analysis (pp. 79-106). Springer, Cham.

[156] Weijters, A. J. M. M., & Ribeiro, J. T. S. (2011, April). Flexible heuristics miner (FHM). In 2011 IEEE symposium on computational intelligence and data mining (CIDM) (pp. 310-317). IEEE.

[157] Wen, L., Wang, J., van der Aalst, W. M., Huang, B., & Sun, J. (2009). A novel approach for process mining based on event types. Journal of Intelligent Information Systems, 32(2), 163-190.

[158] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2016, September). Using life cycle information in process discovery. In International Conference on Business Process Management (pp. 204-217). Springer, Cham.

[159] Zhu, R., Dai, Y., Li, T., Ma, Z., Zheng, M., Tang, Y., ... & Huang, Y. (2019). Automatic real-time mining software process activities from SVN logs using a naive Bayes classifier. IEEE Access, 7, 146403-146415.

[160] Baier, T., Mendling, J., & Weske, M. (2014). Bridging abstraction layers in process mining. Information Systems, 46, 123-139.

[161] Bala, S., Mendling, J., Schimak, M., & Queteschiner, P. (2018, October). Case and activity identification for mining process models from middleware. In IFIP Working Conference on The Practice of Enterprise Modeling (pp. 86-102). Springer, Cham.

[162] Pourmirza, S., Dijkman, R. M., & Grefen, P. W. P. J. (2017). Correlation miner: mining business process models and causal relations without case identifiers. International Journal of Cooperative Information Systems, 26(2), [1742002]. https://doi.org/10.1142/S0218843017420023

[163] Soares, D. C., Santoro, F. M., & Baião, F. A. (2013). Discovering collaborative knowledge-intensive processes through e-mail mining. Journal of Network and Computer Applications, 36(6), 1451-1465.

[164] Wang, D., Ge, J., Hu, H., Luo, B., & Huang, L. (2012). Discovering process models from event multiset. Expert Systems with Applications, 39(15), 11970-11978.

[165] Geng, L., Buffett, S., Hamilton, B., Wang, X., Korba, L., Liu, H., & Wang, Y. (2009, September). Discovering structured event logs from unstructured audit trails for workflow mining. In International Symposium on Methodologies for Intelligent Systems (pp. 442-452). Springer, Berlin, Heidelberg.

[166] Repta, D., & Stanescu, A. M. (2017, May). Holistic Approach for Process Model Discovery. In 2017 21st International Conference on Control Systems and Computer Science (CSCS) (pp. 471-476). IEEE.

[167] Kurniati, A. P., Kusuma, G., & Wisudiawan, G. (2016). Implementing heuristic miner for different types of event logs. vol, 11, 5523-5529.

[168] Sani, M. F., van Zelst, S. J., & van der Aalst, W. M. (2017, September). Improving process discovery results by filtering outliers using conditional behavioural probabilities. In International Conference on Business Process Management (pp. 216-229). Springer, Cham.

[169] Irshad, H., Shafiq, B., Vaidya, J., Bashir, M. A., Shamail, S., & Adam, N. (2015, July). Preserving privacy in collaborative business process composition. In 2015 12th International Joint Conference on e-Business and Telecommunications (ICETE) (Vol. 4, pp. 112-123). IEEE.

[170] Syamsiyah, A., van Dongen, B. F., & van der Aalst, W. M. (2017, September). Recurrent process mining with live event data. In International Conference on Business Process Management (pp. 178-190). Springer, Cham.

[171] Vasilecas, O., Savickas, T., & Lebedys, E. (2014, October). Directed acyclic graph extraction from event logs. In International Conference on Information and Software Technologies (pp. 172-181). Springer, Cham.

[172] Kumar, R., Bhattacharyya, C., & Varshneya, V. (2010, July). Discovering Business Process Model from Unstructured Activity Logs. In 2010 IEEE International Conference on Services Computing (pp. 250-257). IEEE.

[173] Wang, Y., Zacharewicz, G., Traoré, M. K., & Chen, D. (2017, December). A tool for mining discrete event simulation model. In 2017 Winter Simulation Conference (WSC) (pp. 3066-3077). IEEE.

[174] Wang, Y., Zacharewicz, G., Traoré, M. K., & Chen, D. (2018). An integrative approach to simulation model discovery: Combining system theory, process mining and fuzzy logic. Journal of Intelligent & Fuzzy Systems, 34(1), 477-490.

[175] Banziger, R. B., Basukoski, A., & Chaussalet, T. (2018, June). Discovering Business Processes in CRM Systems by leveraging unstructured text data. In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 1571-1577). IEEE.

[176] Jalali, A. (2014, September). Aspect mining in business process management. In International Conference on Business Informatics Research (pp. 246-260). Springer, Cham.

[177] Maggi, F. M., Bose, R. J. C., & van der Aalst, W. M. (2013, June). A knowledge-based integrated approach for discovering and repairing declare maps. In International Conference on Advanced Information Systems Engineering (pp. 433-448). Springer, Berlin, Heidelberg.

[178] Bose, R. J. C., Maggi, F. M., & van der Aalst, W. M. (2013). Enhancing declare maps based on event correlations. In Business process management (pp. 97-112). Springer, Berlin, Heidelberg.

[179] Sturm, C., Schönig, S., & Jablonski, S. (2018). A MapReduce Approach for Mining Multi-Perspective Declarative Process Models. In ICEIS (2) (pp. 585-595).

[180] Leno, V., Dumas, M., Maggi, F. M., La Rosa, M., & Polyvyanyy, A. (2020). Automated discovery of declarative process models with correlated data conditions. Information Systems, 89, 101482.

[181] Maggi, F. M., Dumas, M., García-Bañuelos, L., & Montali, M. (2013). Discovering data-aware declarative process models from event logs. In Business Process Management (pp. 81-96). Springer, Berlin, Heidelberg.

[182] Sturm, C., Fichtner, M., & Schönig, S. (2019). Full support for efficiently mining multi-perspective declarative constraints from process logs. Information, 10(1), 29.

[183] Sutrisnowati, R. A., Yahya, B. N., Bae, H., Pulshashi, I. R., & Adi, T. N. (2017). Scalable indexing algorithm for multi-dimensional time-gap analysis with distributed computing. Procedia Computer Science, 124, 224-231.

[184] Bernardi, M. L., Cimitile, M., & Mercaldo, F. (2018). Cross-organisational process mining in cloud environments. Journal of Information & Knowledge Management, 17(02), 1850014.

[185] Bernardi, M. L., Cimitile, M., & Maggi, F. M. (2014, December). Discovering cross-organizational business rules from the cloud. In 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) (pp. 389-396). IEEE.

[186] Burattin, A., Cimitile, M., & Maggi, F. M. (2014, September). Lights, camera, action! business process movies for online process discovery. In International Conference on Business Process Management (pp. 408-419). Springer, Cham.

[187] Burattin, A., Cimitile, M., Maggi, F. M., & Sperduti, A. (2015). Online discovery of declarative process models from event streams. IEEE Transactions on services computing, 8(6), 833-846.

[188] Di Ciccio, C., Maggi, F. M., & Mendling, J. (2014, September). Discovering target-branched declare constraints. In International Conference on Business Process Management (pp. 34-50). Springer, Cham.

[189] Di Ciccio, C., Maggi, F. M., & Mendling, J. (2016). Efficient discovery of target-branched declare constraints. Information Systems, 56, 258-283.

[190] Kala, T., Maggi, F. M., Di Ciccio, C., & Di Francescomarino, C. (2016, September). Apriori and sequence analysis for discovering declarative process models. In 2016 IEEE 20th International Enterprise Distributed Object Computing Conference (EDOC) (pp. 1-9). IEEE.

[191] Maggi, F. M., Di Ciccio, C., Di Francescomarino, C., & Kala, T. (2018). Parallel algorithms for the automated discovery of declarative process models. Information Systems, 74, 136-152.

[192] Di Ciccio, C., & Mecella, M. (2012, May). Mining constraints for artful processes. In International Conference on Business Information Systems (pp. 11-23). Springer, Berlin, Heidelberg.

[193] Winter, K., Stertz, F., & Rinderle-Ma, S. (2020). Discovering instance and process spanning constraints from process execution logs. Information Systems, 89, 101484.

[194] Di Ciccio, C., & Mecella, M. (2013, April). A two-step fast algorithm for the automated discovery of declarative workflows. In 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) (pp. 135-142). IEEE.

[195] Schönig, S., Rogge-Solti, A., Cabanillas, C., Jablonski, S., & Mendling, J. (2016, June). Efficient and customisable declarative process mining with SQL. In International Conference on Advanced Information Systems Engineering (pp. 290-305). Springer, Cham.

[196] Maggi, F. M., Bose, R. J. C., & van der Aalst, W. M. (2012, June). Efficient discovery of understandable declarative process models from event logs. In International Conference on Advanced Information Systems Engineering (pp. 270-285). Springer, Berlin, Heidelberg.

[197] Räim, M., Di Ciccio, C., Maggi, F. M., Mecella, M., & Mendling, J. (2014, October). Log-based understanding of business processes through temporal logic query checking. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 75-92). Springer, Berlin, Heidelberg.

[198] Ciccio, C. D., & Mecella, M. (2015). On the discovery of declarative control flows for artful processes. ACM Transactions on Management Information Systems (TMIS), 5(4), 1-37.

[199] Maggi, F. M., Mooij, A. J., & Van der Aalst, W. M. (2011, April). User-guided discovery of declarative process models. In 2011 IEEE symposium on computational intelligence and data mining (CIDM) (pp. 192-199). IEEE.

[200] Schunselaar, D. M., Slaats, T., Maggi, F. M., Reijers, H. A., & van der Aalst, W. M. (2018, July). Mining hybrid business process models: A quest for better precision. In International Conference on Business Information Systems (pp. 190-205). Springer, Cham.

[201] Sturm, C., Schönig, S., & Jablonski, S. (2018). A MapReduce Approach for Mining Multi-Perspective Declarative Process Models. In ICEIS (2) (pp. 585-595).

[202] Maggi, F. M., Slaats, T., & Reijers, H. A. (2014, September). The automated discovery of hybrid processes. In International Conference on Business Process Management (pp. 392-399). Springer, Cham.

7 Appendix

License

Non-exclusive licence to reproduce thesis and make thesis public

I, Lekso Migriauli,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

A Systematic Literature Review On Process Discovery Aspects Of Process Mining Applied In Industrial Contexts,

supervised by Fabrizio Maria Maggi and Chiara Di Francescomarino.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Lekso Migriauli

30/07/2021