IJoSI

DOI: 10.6977/IJoSI.202203_7(1).0003
Z. Lamghari / Int. J. Systematic Innovation, 7(1), 35-45 (2022)

Process Mining: Basic Definitions and Concepts

Zineb Lamghari
LRIT associated unit to CNRST (URAC 29), Rabat IT Center, Faculty of Sciences,
Mohammed V University in Rabat, Morocco
Zlamghari.zineb@gmail.com
(Received 4 September 2021; final version received 7 February 2022; accepted 17 February 2022)

Abstract

Business Process Improvement (BPI) paradigm can be implemented on recorded data of real process execution. This is done by analyzing this information and come up with real insights to BP improvement. Even if such model is not available, the presence of a log of activities is very frequent. So, the key idea is that a log can exist even if no process model is present. The spread way of existing BPI methodologies put forward the complexity of their achievement the BP improvement goal. Moreover, they could be driven by many factors. Nonetheless, the common goal is to speed up generating an improved BP.

A recent trending improvement BP method is process mining, compared with existing BPI methodologies, Process Mining had more computer capabilities to implement BP improvements results. However, there are several ambiguities in understanding their general context that must be defined. In this paper, we present basic definitions and notations related to process mining discipline.

Keywords: Process mining, Business process improvement, discovery, conformity, enhancement.

1. Introduction

For some years, the usage of information systems has been rapidly growing in companies of all kinds and sizes. New systems are moving from supporting single functionalities towards a Business Process Management (BPM) orientation (Van der Aalst, 2016; Lamghari et al., 2019).

In this context, activities that companies are required to perform, to complete their own business, are becoming more complex and require the interaction of several persons and heterogeneous systems. A possible approach, to simplify the management of the business, is based on the division of operations in smaller "entities" and on the definition of the required interactions among them.

The term "Business Process" (BP) refers to this set of activities and interactions. As example, we cite the process of handling of a loan application (service), the process of emergency (healthcare), the process of car manufacturing (production), etc. Indeed, several tasks or activities are executed in one instance of such a process. A process instance is commonly denoted as a case, i.e., the activities of the process that operate on the case. Each case of a process has a defined start point and end point.

The seminal articles on business re-engineering (Hammer and Champy, 1993) and (Davenport, 1993) have established the focus on the processes of an organization in management practice. Furthermore, organizations should radically reorganize their work along their value-adding processes. A large body of work, both from industry and from academia, has been organized around the belief that performant processes are the foundation of any successful organization. The basic problem that is being tackled is: How do organizations define and execute performant processes?

This problem has been addressed from various viewpoints and with several approaches. For example, management trends and strategies such as Business Process re-engineering (Elapatha et al., 2020), Lean management (Saxby et al., 2020), Six sigma (Thomas et al., 2009), research fields and methods such as workflow management, and adaptive case management. Moreover, many software systems for BP execution tools have been proposed. For example, Staffware, COSA, YAWL, Bizagi, Bonita, Camunda, jBPM, IBM Business Process Manager, Oracle BPM Suite, etc.

Beyond, BPM can be seen as the umbrella-term that encompasses all those methods that are concerned with the design, enactment, monitoring, and optimization of processes. The main objective of BPM is to align the process with the objectives of the organization. In this sense, each process must be configured, so that the results of the process lead to the achievement of the business goals.

The BPM approach tends to provide more support for various forms of analysis (e.g., simulation) and management support (e.g., monitoring). In many cases, analysts are interested in a real system behaviour, which may be hidden from domain experts and system engineers. To this purpose, most of the software that is used to define and to help companies in executing such processes, typically, leaves a trace of the executable activities. These traces called (Event log). Log consists of the name of the activity and the time the activity is executed; moreover, it is important to note that the traces are grouped in "instances" (can emerged a set of "cases"). Typically, it is necessary to handle several orders at the same time, and therefore the process is required to be concurrently instantiated several times. These instances are identified by a "case identifier" (or "instance id"), which is another field typically included in the log of the traces. Sometimes, especially small, and medium companies do not perform their work according to a formal and explicit BP; instead, they execute their activities with respect to an implicit sorting. Even if such model is not available, the presence of a log of activities is very frequent. So, the key idea is that a log can exist even if no process model is present.

The Business Process Improvement (BPI) paradigm can be implemented on recorded data of real process execution (traces). This is done by analysing this information and come up with real insights to BP improvement. The spread way of existing BPI methodologies put forward the complexity of their achievement the BP improvement goal. Moreover, they could be driven by many factors. Nonetheless, the common goal is to speed up generating an improved BP. In this context, the most trending business process improvement method is PM (Van der Aalst, 2016). However, there are several ambiguities in understanding their general context that must be defined. Therefore, our paper is organized as follows: Section 2 presents existing methodologies related to business process improvement. Section 3 gives a general introduction to the Process Mining field, starting from the PM definition. Section 2 continues with illustrating PM categories and defining the very basic notion of event logs. Section 3 discusses different process models representations. Moreover, a list of process mining algorithms and tools are given (Section 4). The paper finishes with a comparative study between traditional BPIs methodologies and the process mining methodology, showing process mining advantages (Section 5).

2. Process Mining

PM (Van der Aalst, 2016) is a relatively new field incorporating techniques for the discovery, monitoring, and enhancement of real processes by extracting knowledge from the information system event logs. Indeed, PM bridges two different fields: Process Science and Data Science (see Fig. 1). Process Science is a broad area of process modelling, analysis, and optimization. It incorporates Stochastics (analysis of random processes, using Markov chains, queuing networks, and simulation), Optimization (finding the best possible process implementation by applying mathematical optimization techniques), Operations Management & Research (designing and controlling production processes from management and mathematical modelling perspectives), Business Process Management (methods and techniques for the modelling, execution, and enhancement of processes). Business Process Improvement (Six Sigma techniques and Business Process Re-engineering), Process Automation & Workflow Management (tools and methods for the processes execution, including routing and resource allocation), Formal Methods & Concurrency Theory (analysis of process behaviors, using Petri nets, finite state machines, and other formal models). Data Science incorporates all aspects of data analysis and includes Statistics, Algorithms (providing efficient data processing), Data Mining (methods revealing unsuspected relationships in data sets), Machine Learning (techniques for giving computers capability to learn without being explicitly programmed), Predictive Analysis (methods predicting the future trends), Databases (techniques for storing data), Distributed Systems (infrastructure for data analysis), Visualization & Visual Analytics, Business Models & Marketing (techniques for turning data into real value), Behavioural/Social Science (methods for the analysis of human behaviour), Privacy, Security, Law & Ethics (principles protecting individuals from "bad" data science practices).

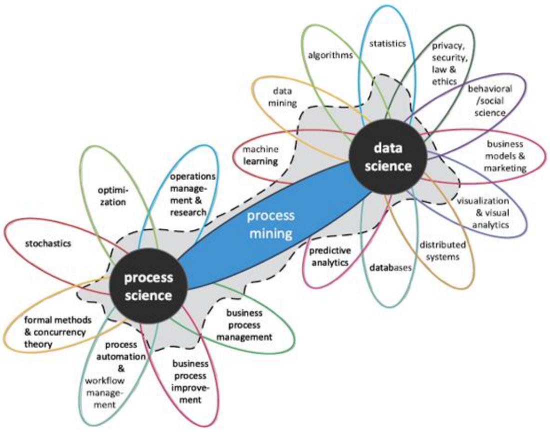

Fig. 1. Overview of process mining and its three types of techniques

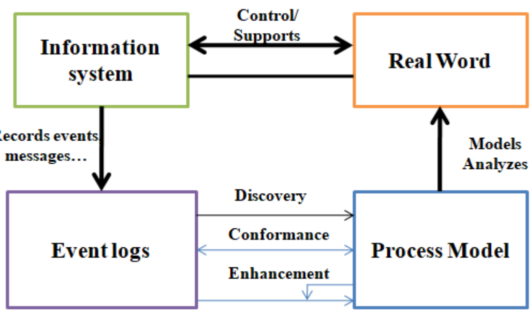

Fig. 2. Overview of process mining and its three types of techniques

2.1 Categories

PM (see Fig. 1) is defined by three categories (Van der Aalst et al., 2011): (1) process discovery, (2) conformance checking, and (3) enhancement.

(1) Process discovery is a challenging task for many reasons. Often event logs are incomplete, i.e., only a fraction of possible behaviors is observed. The other issue is that it can be difficult to uncover the composition of choices, iterations, or parallel executions, represented in the form of a flat event log. It describes process mining only in the offline setting, i.e., only finished process cases are analysed. Generally, process mining is not limited to the offline setting. It also entails methods such as prediction and recommendation based on current process data in an online setting. In the scope of this thesis, we consider the offline and the online setting. The challenge here is the demand of high-quality data and the structured form of resulted processes.

(2) Conformance checking methods find deviations from the expected behaviour. The expected behaviour can be represented in the form of a process model or an event log. One of the main challenges is the computational complexity. Typically, complicated process models and event logs lead to an exponential growth of possible alignments. The other challenge is to provide an intuitive visualization of alignments, helping analysts to reveal important discrepancies. Beyond activity names and timestamps, an event log may contain additional information, such as performers, costs, IP addresses, or other domain specific data.

(3) Enhancement techniques enrich process models with this information. Besides additional attributes taken from the event logs, these could also be results of conformance checking or performance analysis techniques. Model enhancement also considers an event log and a process model as inputs. This means, it is possible to improve an existing process model by looking in the past. Common aspects in model enhancement are time and cost. After discovering a process model from an event log, the discovered process model can be used to analyse for performance indicators, for example average throughput time and costs for improving or re–engineering the process. The bottleneck problem can be identified by analyzing waiting times between activities. After identifying the cause of bottlenecks, the process model can be enhanced at the right places. Enhancing resource performance is one important aspect. A social network in a workplace can be constructed by process discovery. It can give an idea of work collaborations and balance workload to improve resource performance. There is no restricting procedure how a process model can be enhanced. It depends on what problems an organization discovers and how an organisation wants to improve.

2.2 Event Logs

The PM typically assumes that BP execution data are stored as event logs. An event can be considered as the starting point of process mining. The event log structure consists of cases or completed process instances. Each case is made of a sequence of events, called a trace. An event can have any kind of additional attributes (timestamps, cost, resource, etc.) depending on the organization purposes. These additional attributes are important for monitoring the BP improvement. For example, bottlenecks cause that can slow-down the process flow. The event logs notation may depend on the information system treatment or purposes. However, the important point is the quality of these events that can heavily affect the process model representation and by necessity the main business of the organization. Therefore, event logs should be treated as first-class.

a. Notations 1 (Event, Trace and sequence or case): From a mathematical standpoint, each event in an event log is assigned to an activity executed for a singular process instance (one trace). For each trace, all events belonging to that case are ordered in a chronological style (see Figure 2.3.1). In this regard, A = {a1, a2, . . . , an} denotes a finite set, where ai, i:1, 2, ….,n is an activity of a case or sequence of length n. Thus, one case l can be expressed as l=<a1, a2, . . . , an > and < > denotes an empty sequence. The event logs L with n cases and r repetitions can be expressed as L = [$\mathbf{l_1^r}$, $\mathbf{l_2^r}$, …, $\mathbf{l_n^r}$]. For instance, the event logs L=[< a, b, c, d>20 , <a, c, b, d>15] signifies l1 =< a, b, c, d> repeated 20 times, l2 =<a,c,b,d> repeated 15 times, etc.

b. Notations 2 (Noise, Infrequent, incomplete, and chaotic): In a real-life setting, without a-priori knowledge on the process, it is difficult to distinguish between data quality problem: What is considered undesirable behavior depends on the application setting? When looking at the mainstream behavior of the process? then, all infrequent behavior is undesirable. When looking for workaround and divergent process executions, where some infrequent behavior may be desirable.

To explore these deficiencies, we use the following notations: $L = [l_1^r, l_2^r, ..., l_n^r]$ is the event logs, where its composite sequences can be denoted as $l_0, l_1, ..., l_n$ and one sequence or case is expressed as

$$l = \langle a_1, ..., a_n \rangle.$$

1. Noise: entail outliers that were recorded due to errors (Incorrectly logged).
2. $A = \{ a_1 / \ a_1 \neq a_{i,} \quad i = 2,..., n \}$. This point describes events of activities that were executed out of the normal order.
3. Infrequent: low-frequent behaviour. For instance, events recorded due to temporary workarounds, and they are correctly logged. $L = \{min_{l\in L} \text{ where } min_{l\in L} \neq max_{l\in L}\}$
4. Incompleteness: Partial traces. In this paper we mean by incompleteness the problem of missing events. For instance, $U = \{\langle\,\rangle / \,\exists\, i = 1,..., n, \quad a_i \in l\,\}$. To do so, traces are not complete in term of execution, i.e., events must be executed in the normal process but are not observed in the recorded traces).
5. Chaotic activity: can happen anywhere in the process Ch={a_1, a_2, . . . , an / i= 1,…,n }. For example: L=[<a, b, c, d, C_h>,<a, b, c, C_h , d>,< a, b, C_h , c, d>,< a, C_h , b, c, d>,< C_h , a, b, c, d>]. C_h takes any position from a1 to an. Ch is a chaotic activity. We will learn how to filter chaotic activities in the following sections.

3. Process Models

Process models (see Figure 3) are used to visualize, describe, prescribe, and explain the behaviour of processes of an organization for a wide range of objectives such as: communication among stakeholders, process improvement, process management, process automation, and process execution support. Concrete examples are the comparison of the "as-is" and the "to-be" process, documentation for complying with regulatory requirements such as ISO 9001, and the analysis of performance related problems such as bottlenecks and inefficiencies. Depending on the goal of the event logs analysis and on the analyst's personal taste, several ways of process visualization can be used. The most common are Petri Net, Transition Systems, Petri Net and Business Process Management Notation (BPMN).

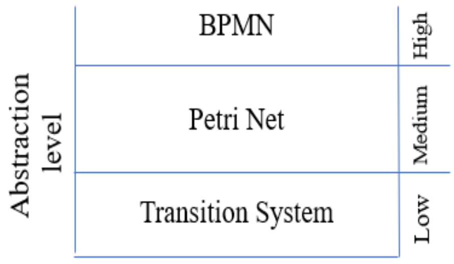

Fig. 3 Process Modelling Notation

The main benefits of adopting a clear business model, through different levels of abstraction, are summarized in the following two points:

(1) It is possible to increase the visibility of the activities, that allows the identification of problems (e.g., bottlenecks) and areas of potential optimization and improvement.
(2) Grouping the activities in "department" and grouping the persons in "roles", in order to better define duties, auditing and assessment activities.

These workflow languages aim at constructing a well-defined and highly automated BP. As a result, processes become more structured. Structured process is referred to rigorously defined process, less complex and with high repetition frequency. The definition of a structured process as given by Devonport is as follows: Structured Process/Lasagna (see part a of Figure 4) is defined as a specific ordering of work activities across time and place, with a beginning, an end, and clearly identified inputs and outputs: a structure for action (Augusto et al.,2018). Therefore, structured processes comprehend those processes whose activities execution consistently follows a predefined process model reference (Werner and Gehrke, 2013). Since a formal representation of these processes can be easily described prior to their execution, tightly framed processes are characterized as fully predictable and repetitive and after their design-time description, they can be repeatedly instantiated at runtime. Examples of this category are production and administrative processes and as well as bank transactions that are executed in an exact sequence to comply with legal norms.

Unstructured process/spaghetti: every instance of the process can be different from another based on the environment, the content and the skills of the people involved. These are always human processes. These processes may have a framework or guideline driving the process but only as a recommendation (Bukhsh et al., 2017). Therefore, unstructured process (see part b of Figure 4) is partially or totally not predefined, adaptable, content–driven and knowledge worker involved. We can then define their characteristics as follows:

- Diversity: processes that can generate a set of execution cases that are structured very differently.
- Knowledge–Intensive: Decisions in an unstructured process that are based on a lot of information, which may be provided from different resources.
- Uncertainty: There is no single answer or end–result in an unstructured process.
- Flexibility: There is no single way to complete an unstructured process. Each next step depends on the previous one and could be completely different every time we run the process.

Moreover, figure 4 illustrates the transformation of a spaghetti process to become more comprehensible as a simplified process model. The operation simplifies process models by keeping high frequent behaviors and filtering out low frequent behaviors. Indeed, executing loosely structured processes generates unstructured behaviors (Taylor, 2014). After mining an unstructured log, a spaghetti-like process can be revealed. "Spaghetti processes" is a metaphor of unstructured ones.

It cannot be assumed that a spaghetti–like process is wrong or that it has a problem caused by process discovery algorithms. It rather means a process model accurately reflects reality. However, spaghetti processes still have issues that are difficult to be analyzed and hard to understand due to its complexity. In this context, it is a very interesting challenge to simplify an unstructured process into a more structured one. The BP characteristics (see Figure 5) can help to determine the system within we will implement our business processes. These systems can be defined as systems with intensive data or with intensive knowledge (see Figure 5). Here, we illustrate the passage between data–intensive system and knowledge–intensive system. Obviously, unstructured processes are composed of a set of structured processes. Indeed, process mining concept consists of treating well-structured processes. For this reason, PM still meaningfully usable for unstructured processes. Consequently, PM can support system with intensive knowledge, like Enterprise Content Management (ECM) and Adaptive Case Management (ACM) system, and systems with intensive data like BPM.

In (Hammer and Champy, 1993; Webber, 2011), the authors focus on how redesigning BPs by presenting four dimensions to quantitatively measure process performance (time, cost, quality and flexibility). This method cannot treat the three first phases (Design, Model, Execute) of the BPM life cycle because it is used after the execution phase. Further, the work in (Davenport, 1993) defines improvement metrics only in the monitor phase by controlling data related to the BP objective. Then, as stated in (Elapatha et al., 2020), the authors present a scientific method to evaluate BPM products by defining a set of criteria for each BPM phase except the optimization phase. Also, as reported in (Saxby et al., 2020) the author focuses on the objectives of performing BPM design. As well as cited in (Thomas et al., 2009), the authors identify the new functional requirements for a Semantic Business Process Management (SBPM) System for each phase of the BPM life cycle and explain the benefits of adopting semantic technologies in SBPM. The authors specify requirements, rather than solutions and metrics. Thus, in (Baluch et al., 2012) the authors propose a method to evaluate and monitor the business process against performance requirements and show the effects of ongoing processes on business goals, in a real-time manner. Additionally, the paper of (Harry, 1998) treats only the implementation phase of the BPM lifecycle. Furthermore, the authors in (De jong et al., 2016) use data warehouse in the BPM life cycle in order to support the tree following phases (Execute, Monitor, Optimize). Last, the work in (Kanji, 1990) considers the whole BPM life cycle by implementing techniques for process mining and intelligent (re)design to support the redesign and diagnosis phases and thus close the BPM life cycle.

4. Process Models

In this section, we resume existing process mining algorithms and tools

4.1 Algorithms

At the heart of a PM discovery technique lays process mining algorithms. These algorithms, often embedded in the process mining software, translate the data from event logs into readable models. Several algorithms are available, each having its own properties concerning the form of input, conversion of data, and form of output. One must pick the right algorithm for a dataset for the right goal and right way of visualization. Much has been written about the mining algorithms. Thus, a process discovery algorithm constructs a generic process model based on event logs. It is an abstracted and general representation of real event logs. Several discovery algorithms are described with basic representation of process models, like alpha algorithm. Other algorithms are representing different levels of abstraction combined with clustering and classification techniques, to model processes from unstructured and complex events. In this regard, we are inspired by (Van der Aalst et al., 2018) to list the following process discovery algorithms:

The first miner developed is the Alpha–algorithm (Sang, 1991). This algorithm is based on eight simple mathematical definitions and visualizes its models in the Petri Net modelling language. Because of its simplicity, it is popular among scholars, but it is unpractical in real-life, because of its difficulty in handling noise, infrequent/incomplete behavior, and complex routing constructs. A second miner is the Heuristic miner (previously called Little Thumb), is better equipped to handle complex routing and it can abstract exceptional behavior and noise, making it suitable for actual logs (Vanden Broucke and Weerdt, 2017). The third algorithm is Fuzzy miner that focusses on unstructured behavior and large event logs. Its output is configurable to reach a desired level of abstraction but can only be visualized in a fuzzy model (Gunther and Van der Aalst, 2007). Fuzzy mining adaptive process simplification based on multi perspective metrics in the international conference on business process management. Springer, Berlin, Heidelberg). Indeed, it deals with process complexity. It highlights significant information and hides less significant activities. In this sense, fuzzy miner simplifies unstructured processes.

The simplification process aims at preserving significant behavior, while less significant but highly correlated behaviors are aggregated into clusters, and less significant or less correlated behaviors are abstracted. Other interesting algorithms are:

(1) Inductive Miner (IM): treats events by grouping them into sub-logs. For each sub–logs, a sub-process is generated. Then, a combination between the resulted sub-processes are released to obtain the generic process model. In this respect, the IM algorithm produces sound models (Bogarin et al., 2018), i.e., less none-conformities detected, and it fits with the majority of present logs. Besides, it cannot identify complex and non-local process control patterns.
(2) Genetic Miner (GM): randomly creates process models from logs. For each process, the precision metric is calculated. Then, sound models are combined based on the mutation operation. The genetic algorithm improves models according to specific objective. The main limitation of this approach is their complexity in discovering and representing process models from real data sets (Vanden Broucke and Weerdt, 2017). State Based Regions (SBR): generates a Petri net from a Transitions System (ST) based on specific abstractions, such as: Set, Multi-Set, Sequence, and other types of abstractions, in which each state of the ST can be represented by a complete or partial trace. This algorithm ensures the fitness metric, as well as the identification of complex control structures. On the other hand, SBR is unable to process incomplete and noisy logs (Van der Werf et al., 2008).
(3) Language Based Regions (LBR): The main idea of this algorithm is to find places based on the language process. All candidate places correspond to a language region. The discovered Petri net will be obtained by adding a place for each positive solution based on linear resolution. Each solution is represented as a triplet (A; B; C), where A is the set of inputs arcs, B is the set of outputs arcs and C is the number of tokens in the square. Indeed, the LBR algorithm uses properties derived from logs (causal relationships), to determine the final model by describing different places. Unfortunately, this algorithm is unable to process incomplete and noisy logs (Van der Aalst et al., 2018).

4.2 Tools

Currently, there is a wide range of research and commercial tools available in PM area: ProM[1], Disco[2], ARIS Process Performance Manager[3], QPR Process Mining[4], ProcessGold[5], Celonis[6], Minit[7], and myInvenio[8]. Indeed, several process mining tools are available. Choice can be based on a specific needed set of functionalities, supported data formats, but also costs. Finding the right tool can be difficult since no comprehensive comparison exists.

Table 1 shows a list of process mining tools. Overall, three earnings models can be distinguished. First the common licencing structure where an organization can buy the tool for a certain period or indefinitely. The tool can be sold in combination with or without support for implementation or analysis. This is the case with, for example, ARIS BPM (Process Performance Manager), Celonis Process Mining, and Disco. Another earnings model is offering process mining as a service (PMaaS). This is provided by Icris and Coney. And lastly, several open-source tools are available. Most famous example is ProM, but also Apromore is popular. Based on the number of academic publications on the topic, open-source process mining platform ProM seems to be the most popular. ProM is an extensible framework that runs on Java and obtains its functionality by a wide variety of plug-ins. Because it is an independent platform and is developed by process mining "creator" Will van der Aalst, it is popular among scholars performing applied research. There are over 1500 free plug-ins available, each with different functionalities and options (Van der Aalst, 2016). For example, the use of different miners (heuristic, alpha-algorithm, and fuzzy), sorts of output (Petri Net and BPMN), and types of analysis (process discovery, dotted chart, and social networks). However, the academic character makes it difficult to use. Manuals and instructions are missing, and support can only be found in its community of volunteers. It seems that only few commercial organizations use this tool and even if they do so, they mostly use it to learn the concept of process

[1] http://www.promtools.org/doku.php

[2] https://fluxicon.com/disco/
[3] http://www2.softwareag.com/
[4] https://www.qpr.com/solutions/process-mining
[5] http://www.processgold.com/en
[6] https://celonis.com/
[7] https://www.minit.io/
[8] https://www.my-invenio.com/process-mining-vision/

mining before buying more user-friendly tool.

A more user-friendly process mining tool package is Disco, Developed by Fluxicon. Disco lacks some functionalities when compared to ProM, but distinguishes itself with a fast, well documented, and clear interface. Very limited knowledge on process mining is required to perform an analysis. But the lack of real-time connections to databases makes it less useful for large companies. It seems logical that Fluxicon focusses on small and medium-sized enterprises.

A third tool package, used by MoD, is ARIS Process Performance Manager. This package contains three components: the administrator's section, the business analyst's section, and the dashboard.

The section for the administrator is used to load the data. This can be a single file, but also a connection to a database. A good example is SAP. With the right connector, the SAP databases can be periodically and automatically loaded into the ARIS databases. Initializing this connector will cost some effort but can be a good investment. The business analyst's section of the tool is used for in-depth analysis. With the use of filters, selections of the dataset can be made. Analysts can use several techniques and models to answer their process related questions. The steps of this process (the query) can be saved, so when the data are refreshed, the analysis can be updated.

The last section, the dashboard, is meant for tracking the organization's process performance. The analyst can develop certain queries. For example, the average lead time of preventive maintenance of a specific weapon system, and this can be loaded into the process-centric dashboard as a Key Performance Indicator (KPI). The dashboard periodically collects recent data from the ERP and the process owner can follow the progression. Based on the average process and on its excesses, the process owner can decide to intervene. Also, the (Kebede, 2015) developed a model to compare ProM, Disco, and Celonis on fifteen characteristics. The model was updated to the lasted tool versions and Celonis was replaced for ARIS PPM.

Table 1. Exhaustive list of process mining developers

Start	Software name	Software Developper	Country
2005	ProM	The Process Mining Group	The Netherlands
2007	ARIS Process Performance Manager	Software AG	Germany
2007	Interstage Automated Process Discovery	Fujitsu, Ltd.	Japan
2008	StereoLOGIC Discovery Analyst	StereoLOGIC	The United States
2009	The Process Mining Factory	Icris	The Netherlands
2010	Apromore	The Apromore Initiative	Australia
2011	Celonis Process Mining	Celonis GmbH	Germany
2011	Perceptive Process Mining	Perceptive Process Mining	The United States
2012	Disco	Flusicon	The Netherlands
2012	QPR Process Analyzer	QPR	Finland
2012	Process Mining Solution	Coney	The Netherlands

2013	SNP Business Process Analysis	SNP Schneider-Neureither & Partner AG	Germany
2013	minit	Gradient ECM	The United States
2015	myInvenio	Cognitive Technology Ltd.	Malta
2015	XMAnalyzer	XMPro	The United States
2016	Lana	Lana Labs GmbH	Germany
2017	ProcessGold	ProcessGold International B.V.	The Netherlands
2018	Kofax insights	Kofax INC.	The United States
2018	Appnomic OpsOne	Appnomic self-healing enterprise	India
2019	MPM process mining	Mehrwerk GmbH	The Netherlands
2020	EverFlow	Icaro Tech	Brazil
2021	UiPath Process Mining	UiPath	Romania

Table 2. Process mining compared to traditional BPI methodologies

Concepts	Process Mining (Mature version)
First mentioned	2015
Origin	The rise of big data and accessibility of computing power
Align	Align de facto with de jure process models
Process view	Discovering, conforming, and enhancing business processes
Involvement	Multidisciplinary team
Methodology	Plan, extract, process data, mine and analyze event data, evaluation, and process improvement & support in a team
Primary Effects	Gain quantitative and factual knowledge about processes
Secondary Effects	Improvements can be monitored and verified
Criticism	Demanding high quality data and structured processes

5. Discussion and Conclusion

A basic discussion of several traditional BPI methodologies has already been given in section 1. In this research, we compare process mining with these traditional methodologies. On one hand, they find their origin in Lean and while throughout the years many different BPI approaches arose, the boundaries between these BPI approaches remain vague. In this regard, many approaches have been combined to form new BPI approaches and users often interchange the terms. As a science, this makes it difficult to investigate their characteristics. An overview is presented to order eight widely used BPI approaches on nine properties.

This overview contributes to the knowledge on BPI. However, this is only the tip of the iceberg. Scholars and management consultants frequently come up with new BPI approaches and combine them till distinction is long gone. For example, next to Lean Six Sigma and Lean MRO (Maintenance Repair and Overhaul) the research shows us that the variations Lean Start-up, Lean Manufacturing, Lean Management, Lean Thinking, Lean Enterprise, and Lean Maintenance also are being used. These variations could be completely new methodologies, slightly adjusted methodologies, or identical to Lean as described in this chapter.

On the other hand, all discussed BPI approaches have a statistical or analytical background, but process mining excels in its ability to automatically convert data into organized information. It is hypothesized that PM is a mature BPI approach. In its short history, process mining has made an impressive development. Every year, more process mining tool is developed, papers on the topic are published, courses in process mining are given, and case studies are conducted (Gonella, 2017). Since process mining is a data driven activity, and with data storage becoming cheaper and cheaper and initiatives like the

Internet of Things (IoT) boosting data production, new possibilities do arise.

Mechanics can, for example, enter their activity data in the ERP with wireless tablets, giving real time analysis possibilities. Combining this with Artificial Intelligence (AI) and Machine Learning (ML) can offer even bigger opportunities. AI can find deviating process instances and even suggest improvements without human intervention. PM provides data on all activities of a process: its throughput, lead times, its delays, etc. These data can be used for building an accurate model in simulation tools. By reasoning, but also by trial-and-error, elements in the simulation can be changed till the model cannot be improved anymore. The changes can then be applied in the real world. Integrating process mining tools with simulation tool can create significant opportunities, and from the managerial standpoint, PM accumulates teams experiments to produce more significant results.

To conclude, all discussed BPIs have a statistical or analytical background, but process mining excels in its ability to automatically convert data into organized information. Compared with existing BPI methodologies, process mining had more computer capabilities to implement BP improvements results. However, there are several drawbacks that must be addressed. Indeed, these drawbacks will be discussed in the following chapter. They mainly concern event logs quality and business process structures.

References

Augusto, A. Conforti, R. Dumas, M. La Rosa, M., Maggi, F.M., & Marrella, A. (2018). Automated discovery of process models from event logs: Review and benchmark, IEEE Transactions on Knowledge and Data Engineering, 31, 4, 686-705.

Baluch, N.H., Che Sobry, A., & S. Mohtar, S. (2012). TPM and lean maintenance: A critical review. Interdisciplinary Journal of Contemporary Research in Business (IJCRB), 4, 2, 850-857.

Bogarin Vega A. , R. Menendez, & Romero, C. (2018). Discovering learning processes using inductive miner: A case study with learning management systems, Psicothema, 30, 3, 322-329.

Bukhsh, Z.A., Van Sinderen, M., K. Sikkel, K., & Quartel, D.A. (2017). Understanding Modeling Requirements of Unstructured Business Processes. In 14th International Conference on e-Business, King Juan Carlos University, Madrid, Spain.

Davenport, T.H. (1993). Process innovation: reengineering work through information technology, Harvard Business Press.

De Jong, S.J., Wouter, W, & Van Blokland, B. (2016). Measuring lean implementation for maintenance service companies. International Journal of Lean Six Sigma, 7, 1, 35-61.

Elapatha, V., Wijeratne, J., & Shahzadah, N. (2020). An analysis of the implementation of business process re-engineering in public services. Journal of Open Innovation: Technology, Market, and Complexity, 6, 4, 114.

Graafmans, T., Turetken, O., Poppelaars, H., & Fahland, D. (2021). Process mining for six sigma. Business & Information Systems Engineering, 63, 3, 277-300.

Gonella, P. (2017). Process mining: A database of applications. HSPI Management.

Gunther, C.W., & Van der Aalst, W.M.P. (2007). Fuzzy mining adaptive process simplification based on multi perspective metrics. International conference on business process management, Springer, Berlin, Heidelberg.

Hammer, M., & Champy, J. (1993). Reengineering the Corporation: A Manifesto for Business Revolution. Paperback edition, HarperBusiness, New York.

Harry, M.J. (1998). Six sigma: A breakthrough strategy for profitability, Quality Progress, 31, 5, 60-64.

Kanji, G.K. (1990). Total quality management: the second industrial revolution, Total Quality Management, 1, 1, 3-12.

Kebede, M. (2015). Comparative Evaluation of Process Mining Tools. P.hd thesis, University of Tartu, Estonia.

Lamghari, Z., Radgui, M., Saidi, R, & Rahmani, M.D. (2019). Passage challenges from data-intensive system to knowledge–intensive system related to process mining field, Proceedings of the ArabWIC 6th Annual International Conference Research Track-ACM.

Saxby, R., Cano-Kourouklis, M., & Viza, E. (2020). An initial assessment of lean management methods for industry 4.0. The TQM Journal, 32, 4, 587-601.

Thomas, A. Barton, A., & Chuke-Okafor, C. (2009). Applying lean six sigma in a small engineering company–a model for change. Journal of Manufacturing Technology Management, 20, 1, 113-129.

Sang, R. (1991). The finite element method for elliptic problems. Journal of Elasticity, 38, 2, 209-218.

Taylor, J. (2014). Decision management solutions. Retrieved 2014 from http://www.decisionmanagementsolutions.com

Tax, N., Sidorova, N., Haakma, R., & van der Aalst, W.M.P. Mining process model descriptions of daily life through event abstraction. In Intelligent systems and applications, (2017), Springer.

Vanden Broucke, S.K., & J. Weerdt, (2017). Fodina: A robust and flexible heuristic process discovery technique. Decision support systems, 100, 109-118, 2017.

Van der Werf, J.M.E., Van Dongen, B.F., Hurkens, C.A., & Serebrenik,A. (2008). Thriving on Adaptability: Best Practices for Knowledge Workers. Springer.

Van der Aalst, W.M.P., Rubin, V., Verbeek, H.M.W., Van Dongen, B.P., Kindler, V., & Gunther, C.W. (2010). Process mining: a two-step approach to balance between underfitting and overfitting. Software & Systems Modeling, 9, 1, 87.

Van der Aalst, W.M.P., Adriansyah, A., De Medeiros, A., Arcieri, F.T. Baier, Blickle, T., & Chandra, B. (2011). Pocess Mining Manifesto. The international Conference on Business Process Management, Springer.

Van der Aalst, W.M.P. (2016). Data science in action: Process discovery, 2nd ed. Springer, Berlin, Heidelberg.

van Zelst, S. J., Mannhardt, F., de Leoni, M., & Koschmider, A. Event abstraction in process mining: literature review and taxonomy. Granular Computing, 6, 3, 719-736, (2021).

Webber, L., & Wallace, M. (2011). Quality control for dummies. John Wiley & Sons.

Werner, M., & Gehrke, N. (2013). Process mining, die Zeitschrift fur den Wirtschaftsstudenten, 7, 13, 1-16.

Biography:

Zineb Lamghari, Doctor in software engineering. received her technical university degree in software engineering at the Higher School of Technical Education, Mohammed V-Souissi University, Morocco, and her license degree in information systems management from polydisciplinary faculty then her master's degree in computer system and network from the faculty of Technical Sciences, Abdelmalek Essaadi University, Morocco. She prepares her PhD in business process improvement in the Computer Science and Telecommunication Research Laboratory (LRIT) at the faculty of sciences, Mohammed V University. Her interests cover mainly business process management, software engineering and process mining technique.