

# A Survey on Process Mining

Atuba Zooni N<sup>1</sup> Dr. B.M. Vidyavati<sup>2</sup>

<sup>1</sup>M.Tech. Student <sup>2</sup>Professor

<sup>1</sup>Department of Computer Science & Engineering

<sup>1,2</sup>BITM, Ballari

**Abstract**— Process mining is the process management system which deals with massive amount of data, documents and files i.e., in terra bytes. Process mining improves business processes by keeping track of all the data logs, event logs and transaction logs in business environment. It can also be applied to software development and information system- a model can be derived out of the software development process. The main aim of process mining is extraction of event logs, and to do this a process mining framework is used for obtaining software process models as well as for analyzing and optimizing the same. A process mining framework called ProM framework was developed by Wil M.P. van der Aalst and his fellow researchers at Eindhoven University of Technology. ProM is an open source toolkit and it helps extraction of event logs. Process mining includes various challenges which opens a door for lot of research and development for the enhancement of this technology.

**Key words:** Control-Flow Discovery, Process Mining

## I. INTRODUCTION

Process mining is young research area that has been evolved in the late 1990s in the context of analyzing software engineering processes by Cook and Wolf in the late 1990s (Cook and Wolf 1998). Agrawal and Gunopulos (Agrawal et al. 1998) and Herbst and Karagiannis (Herbst and Karagiannis 1998) introduced process mining to the context of workflow management<sup>[1]</sup>. Process mining is applicable in many areas but initially it had been developed to serve the business performance in an organization and it was called as Business Process Mining (BPM). Process mining basically extracts knowledge from the event logs that are readily available from the information systems. The main aim of process mining is extraction of event logs from the readily available in information systems, and to do this a process mining framework is used for obtaining software process models as well as for analyzing and optimizing the same. Process mining is process centric, truly intelligent and fact-based. The whole idea of mining revolves around the event logs, each event is referred as an activity.

The goal of process mining is to extract information about processes from transaction logs. It assumes that it is possible to record events such that<sup>[2]</sup>

- Each event refers to an activity (i.e., a well-defined step in the process)
- Each event refers to a case (i.e., a process instance)
- Each event can have a performer also referred to as originator (the actor executing or initiating the activity)
- Events have a timestamp and are totally ordered.

Systems such as Enterprise Resource Planning (ERP) systems or Customer Relationship Management (CRM) produce large amounts of data, which can be analyzed using various techniques and tools such as

Business Intelligence (BI), Online Analytical Processing (OLAP) and Data Mining. This whole process, known as the BPM lifecycle, is depicted in Figure 1. The data collected throughout the BPM lifecycle can be used for performance analysis and redesign, but also for detecting (intentionally) deviating behavior.

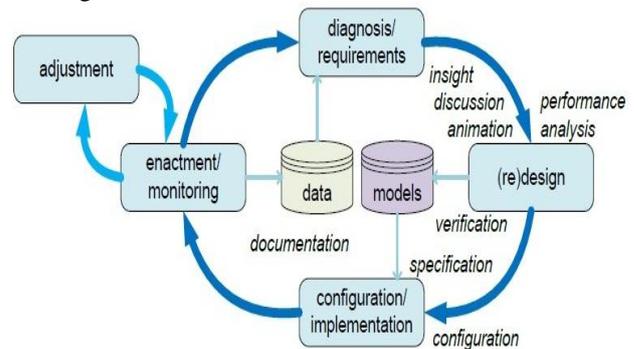


Fig 1: Process mining overview. Taken from (van der Aalst, 2011).

Figure 1 illustrates how a process model can be adjusted, monitored, configures or designed/redesigned in its lifecycle. The cycle consists of two databases, data and model. The former consists of documents and the latter consists of models that have already been processed.

A process model can enhance its performance analysis after performing certain diagnosis on the already processed models, and this shall include insight, discussion, animation etc., and with new requirements this shall be modified. The (re)design phase performs some verification and specification tasks to get the configuration done. Similarly all the other phases perform their tasks efficiently to get the optimal process model. To do this process mining has several characteristics which helps process models and event logs enhance its processing.

## II. CHARACTERISTICS OF PROCESS MINING

The below characteristics has been suggested in the process mining manifesto which has a huge impact on the modeling and processing of the models and their event logs. The characteristics<sup>[3]</sup> are as follows:

### A. Process mining is not limited to control-flow discovery:

Control-flow discovery is the exciting part of process mining. The event logs help in discovering of process models which indeed fuels the imagination of both practitioners and academics. Discovery belongs to one of the three basic forms of process mining viz., discovery, conformance, and enhancement. Process mining is not limited to the scope of control-flow discovery. There are other perspectives that also play an important role such as time, case and organizational perspectives.

**B. Process mining is not just a specific type of data mining:**

Process mining can be represented as the bridge or a missing link between traditional model-driven BPM, analysis, and data mining. The techniques in data mining are not process-centric at all. This includes most of the techniques in data mining. Data mining structures such as association rules and decision trees are incomparable to process model which potentially exhibiting concurrency. This leads to the generation of completely new types of algorithms and representations.

**C. Process mining is not limited to offline analysis:**

The techniques in process mining are used to extract knowledge from historical event data. The results obtained from the “post mortem” data can be applied to running cases. For example, the completion of a partially handled customer order can be predicted using a discovered process model.

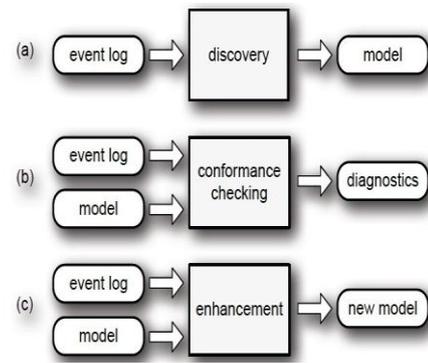


Fig. 3.1: Three basic types of process mining [3]

**III. TYPES OF PROCESS MINING**

As per the process mining manifesto there are three basic types of process mining. They are as follows:

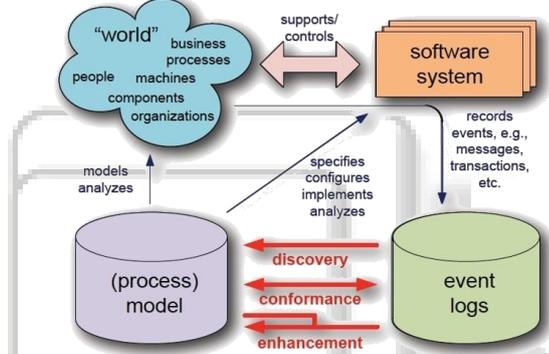


Fig 3: Positioning of the three types of process mining [3]

**A. Discovery:**

Discovery or process discovery is the most prominent technique in process mining. The discovery technique takes an event log as input and produces a model without using any a-priori information. Surprisingly for many organizations, based on example executions in event logs the existing techniques are able to discover real processes.

**B. Conformance Checking:**

The process mining manifesto defines conformance checking as: conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. In conformance checking, an existing process model is compared with an event log of the same process. Different types of models can be considered in conformance checking like: procedural models, organizational models, declarative models, business rules/policies, laws etc.

**C. Enhancement:**

The idea here is, using the information about the actual processes recorded in some event log. The existing process model can be extended or improved. The main aim is to change or extend the a-priori model. For example, by using timestamps in the event log one can extend the model to show bottlenecks service levels, throughput times, and frequencies.

**IV. THE PROM FRAMEWORK**

The ProM framework is the standard framework for process mining that is currently available in the market. It is open source and has various versions to it. ProM supports multiple formats and multiple languages, e.g., Petri nets, EPCs, Social Networks, etc. The plug-ins can be used in several ways and combined to be applied in real-life situations. Adding a new plug-in is easy, i.e., new mining plug-ins can be added without re-compiling the source code.

**A. Architecture of ProM Framework**

The basis for all process mining techniques is a process log. Such a log is a file generated by some information system, with information about the execution of a process. Since each information system has its own format for storing log files, we have developed a generic XML format for the ProM framework to store a log in. This format was based on a thorough comparison of the input needs of various existing (ad-hoc) process mining tools and the information typically contained in an audit trail or transaction log of some complex information system (e.g., an ERP or a WFM system).

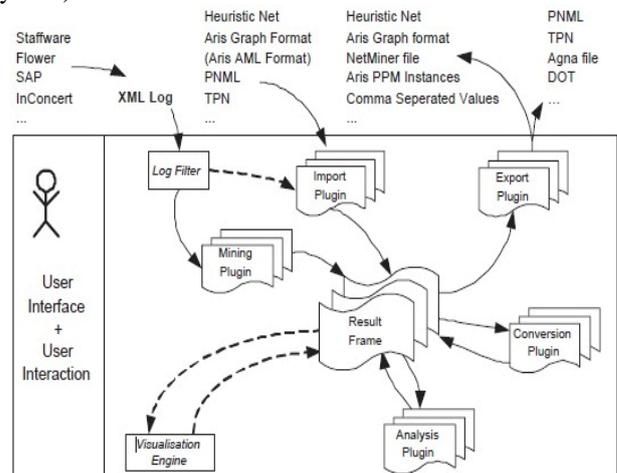


Fig 4.1: Overview of the ProM Framework [2]

Another important feature of the ProM framework is that it allows for interaction between a large number of so-called plug-ins. A plug-in is basically the implementation of an algorithm that is of some use in the process mining area, where the implementation agrees with the framework. Such plug-ins can be added to the entire framework with relative ease: Once the plug-in is ready it can be added to the framework by adding its name to some ini-file. Note that there is no need to modify the ProM framework (e.g., recompiling the code) when adding new plug-ins, i.e., it is a truly “pluggable” environment.

The Mining plug-ins does the actual mining and the result is stored as a Frame. These frames can be used for visualization, e.g., displaying a Petri net, an EPC or a Social network, or further analysis or conversion. The Analysis plug-ins take a mining result and analyze it, e.g., calculating a place invariant for a resulting Petri net. The Conversion plug-ins take a mining result and transform it into another format, e.g., transforming an EPC into a Petri net or vice versa.

### B. Plug-ins in the ProM Framework

Currently, there are nine export plug-ins, four import plug-ins, seven analysis plug-ins, and three conversion plug-ins. There are five kinds of general plug-ins<sup>[2]</sup> available:

- “Mining plug-ins” which implement some mining algorithm, e.g., mining algorithms that construct a Petri net based on some event log.
- “Export plug-ins” which implement some “save as” functionality for some objects (such as graphs). For example, there are plug-ins to save EPCs, Petri nets (e.g., in PNML format), spreadsheets, etc.
- “Import plug-ins” which implement an “open” functionality for exported objects, e.g., load instance-EPCs from ARIS PPM.
- “Analysis plug-ins” which typically implement some property analysis on some mining result. For example, for Petri nets there is a plug-in which constructs place invariants, transition invariants, and a cover-ability graph. However, there are also analysis plug-ins to compare a log and a model (i.e., conformance testing) or a log and an LTL formula.
- “Conversion plug-ins” which implement conversions between different data formats, e.g., from EPCs to Petri nets.

The plug-ins that are available for the three perspectives of process mining are as follows:

For the “process perspective”, four plug-ins are available:

- $\alpha$ -algorithm which implements the  $\alpha$ -algorithm and its extensions as developed by the authors. The  $\alpha$ -algorithm constructs a Petri net which models the process recorded in the log.
- Tshinghua- $\alpha$  algorithm which uses timestamps in the log files to construct a Petri net. It is related to the  $\alpha$  algorithm, but uses a different approach. It is interesting to note that this mining plug-in was the first plug-in developed by researchers outside of our research group. Researchers from Tshinghua University in China (Jianmin Wang and Wen Lijie)

were able to develop and integrate this plug-in without any help or changes to the framework.

- Genetic algorithm which uses genetic algorithms to tackle possible noise in the log file. Its output format is a heuristics net (which can be converted into an EPC or a Petri net).
- Multi-phase mining which implements a series of process mining algorithms that use instance graphs (comparable to runs) as an intermediate format. The two-phase approach resembles the aggregation process in Aris PPM.

For the “organizational perspective”, one plug-in is available:

- Social network miner which uses the log file to determine a social network of people. It requires the log file to contain the Originator element.

Finally, for the “case perspective”, also one plug-in is available:

- Case data extraction which can be used for interfacing with a number of standard knowledge discovering tools, e.g., Viscovery and SPSS AnswerTree.

### V. CHALLENGES IN PROCESS MINING

The process mining manifesto has come up with a list of challenges that the organizations face while managing the non-trivial operations. The manifesto describes the functioning of process mining as: in process mining, on one hand, there is an incredible growth of event data; and on the other hand, processes and information need to be aligned perfectly in order to meet requirements related to compliance, efficiently, and customer service.

The challenges of process mining are not complete, because over the time, new challenges may emerge or the existing challenges may be disappeared as new researches may provide many advances in process mining. Some of the challenges are as follows:

The major challenges include: Finding, merging, and cleaning event data. Improving the representational bias used for process discovery, Balancing between quality criteria such as fitness, simplicity, precision, and generalization, Providing operational support. Combining process mining with other types of analysis, Improving usability for non-experts, Improving understandability for non-experts, Dealing with concept drift, Dealing with complex event logs having diverse characteristics, Creating representative benchmarks, Cross-organizational mining.

Process mining also includes other challenges<sup>[1]</sup> such as noise and incompleteness, competing model quality criteria, event log quality and labeling, and complex process models. The organizations strive to overcome all these challenges to have a better system running, which can provide the best outcomes for all the tasks.

### VI. CONCLUSION

Process mining an emerging technology and many innovations are being done around the globe with respect to this technology. Earlier process mining was meant exclusively for business processes but lately it is being implemented in every area and various researches are being conducted for the same. The ProM framework is the

standard tool for the process mining to be implemented and has plugins for every task the organization needs. The challenges in process mining are evident and many researchers are trying to minimize the effects of each. Furthermore, process mining is a boon to business organizations, which can enhance their businesses by making use of the techniques in process mining.

#### REFERENCES

- [1] Prof. Dr. Nick Gehrke and Michael Warner: "Process Mining article", Dipl.-Wirt.-Inf. WISU-die Zeitschrift für den Wirtschaftsstudenten, 2013.
- [2] B.F. van Dongen, A.K.A. de Medeiros, H.M.W. Verbeek, A.J.M.M. Weijters, and W.M.P. van der Aalst: "The ProM framework: A new era in process mining tool support", Springer-Verlag Berlin, Heidelberg Volume 3536, 2004-2011.
- [3] Process mining manifesto: Official Website: [www.processmining.org](http://www.processmining.org)
- [4] Vladimir Rubin, Christian W. Gunther, Wil M.P. van der Aalst, Ekkart Kindler, Boudewijn F. van Dongen, and Wilhelm Schafer: "Process Mining Framework for Software Processes", June 2011.
- [5] Wil M.P. van der Aalst: "Process Mining book-Discovery, Conformance and Enhancement of Business Process", Springer Publications, April 5, 2011.
- [6] H.M.W. Verbeek, B.F. van Dongen J. Mendling, Wil M.P. van der Aalst: "Interoperability in ProM Framework", 18th International Conference on Advanced Information Systems Engineering, pages 619-630, 2010.

