International Journal of Engineering and Techniques - Volume 1 Issue 4, July - Aug 2015

RESEARCH ARTICLE OPEN ACCESS

A review paper on Process Mining

Mekhala [1]

Roorkee College of Engineering, Roorkee.

Abstract:

This study aims to enlighten the researchers about the details of process mining. As process mining is a new research area, it includes process modelling and process analysis, as well as business intelligence and data mining. Also it is used as a tool that gives information about procedures. In this paper classification of process mining techniques, different process mining algorithms, challenges and area of application have been explained.Therefore, it was concluded that process mining can be a useful technique with faster results and ability to check conformance and compliance.

I. INTRODUCTION

The increasing demand to learn more about how their process operates in the real world in the companies have developed and there is increase in the use of process mining techniques. Process mining addresses the problem that most 'process owners' have very limited information about what is actually happening in their organisation. In practice, there is often a significant gap between what is predefined or supposed to happen, and what actually happens. The aim of Process mining is to allow for the analysis of business process based on event logs. It is a relatively new and emerging research area dealing with the process modelling and process analysis, as well as business intelligence and data mining.

There are two reasons that confirm the efficiency of the mining process. Primary, it is used as a tool that gives information about how people and procedures really work. Since SAP logs all transactions that cover the people and procedures it is a good example for this. Secondary, process mining is a beneficial tool to compare predefined processes and the actual process. The various data mining techniques such as classification, association, clustering are mostly used to analyze a distinct step in the overall business process but cannot be applied to understand and analyze a process as a whole (Dunham et al.,2002). It extracts knowledge from event logs.

An Event can be defined as an activity corresponding to the starting point of the process mining. A sequential relation between events is required in process mining techniques. Each activity is a unique process instance, i.e. it belongs to a specific event. Also, additional information such as the source initiating or carrying out an activity (a person or a device), the occurrence and ending time of events (may be activity-based), or data elements recorded with the incident (such as the size of an order), which is required in order to create a realistic model.

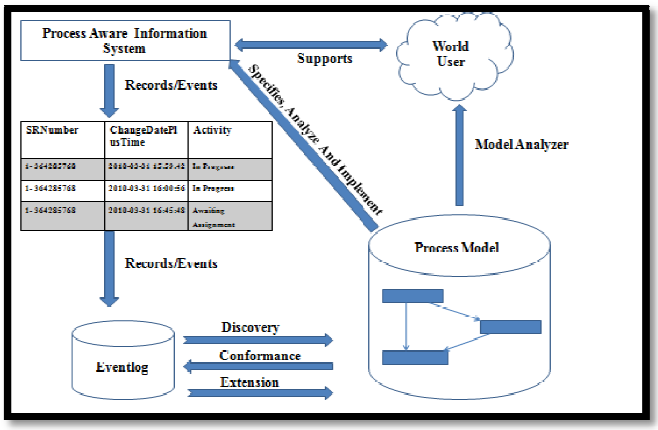

Figure 1.1: Types of Process Mining Techniques

2. Process Mining Techniques

As can be seen from Figure 3, there are basically three types of process mining techniques

2.1. Process Discovery:

This technique take no a-priori process schema particularly based on an event log some schema is constructed. For example, using the alpha algorithm, a process schema can be discovered based on low-level events (van de Aalst et al., 2004).

2.2. Process Conformance:

This technique takes an existing process schema. This schema is used as a reference and check if reality conforms to the schema. For example, there may be a process schema indicating that purchase order of more than one thousands Rupees required two checks. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations (Rozinat and van der Aalst, 2006a).

2.3. Process Enhancement:

This technique takes the existing process schema. This schema is extended with a new aspect or perspective especially the goal is not check conformance but tries to extract new information from it. For example in the detection of data dependencies that affect the routing of a case and adding this information to the model in the form of decision rules (Rozinat and van der Aalst, 2006b).

3. Characteristics of Process mining:

3.1 Process mining is not limited to control-flow discovery

The discovery of process models from event logs fuels the imagination of both practitioners and academics. Therefore, control-flow discovery is often seen as the most exciting part of process mining.

3.2. Process mining is not just a specific type of data mining

Process mining can be seen as the "missing link" between data mining and traditional model driven BPM. Most data mining techniques are not process-centric at all. Process models potentially exhibiting concurrency are incomparable to simple data mining structures such as decision trees and association rules. Therefore, completely new types of representations and algorithms are needed.

3.3. Process mining is not limited to offline analysis

Process mining techniques extract knowledge from historical event data. Despite of the fact "post mortem" data is used; the results can be applied to running cases.

4. Types of process mining algorithms

There are many mining technique to meet process mining with different requirements as alpha-miner, alpha++-miner, heuristic miner and so on. Some of them are discussed below:-

4.1Alpha Algorithm

This algorithm is a relatively intuitive and simple based on dependency relation between events which required ideal event logs without noise. It was one of the first algorithms that are able to deal with concurrency. It takes an event log as input and calculates the ordering relation of the events contained in the log. It has some problems with noise and frequency, and its results are not very world-realistic therefore, this algorithm is not recommended for real logs.

4.2Heuristic Miner

This algorithm focuses on the control flow perspective and creates a model in Heuristics Nets format for a given event log. HM also used determistic algorithm but it extends alpha algorithm by considering the frequency of trace in the log that solve the noise problem by expressing the number of connections between different tasks in the event log. As HM is based on frequency patterns it let the user stay on the main behaviour of event logs.

The Heuristics Miner uses frequencies to calculate a Dependency Measure that which indicate the strength of the causal relation between a pair of activities. Then, the construction of the input and output expressions are taken place for each activity and at last search for long distance dependency relations held.

Steps of Heuristic Miner algorithm

1. Read a log
2. Get the set of tasks
3. Infer the ordering relations based on their frequencies
4. Build the net based on inferred relations
5. Output the net

4.3Genetic Miner

This algorithm uses an evolutionary approach that mimics the process of natural evolution. Despite of the fact the algorithm can mine process models that

might contain all the common structural constructs and can handle noise; it can take a large amount of computational time.

Genetic mining algorithms follow four steps: **initialization, selection, reproduction and termination**.

The goal of using genetic algorithms is to tackle problems such as duplicate activities, hidden activities, non-free choice constructs, noise, and incompleteness, i.e., overcome the problems of some of the traditional approaches.

Main steps of our genetic algorithm are:

1. Read the event log.
2. Calculate dependency relations among activities. Build the initial population.
3. Calculate individuals‟ fitness.
4. Stop and return the fittest individuals?
5. Create next population by using the genetic operators .

5 Software of Process Mining

A list of existing process mining software was gathered by informing existing literature and informal research on the internet (Tiwari, Turner, & Majeed, 2008; W. Van Der Aalst et al., 2011), see table 1.

Name	Company	Country
ProM	The Process Mining Group, Eindhoven Technical University	The Netherlands
Disco	Fluxicon	The Netherlands
Celonis Discovery	Celonis	Germany
Perceptive Process Mining	Perceptive Software, Lexmark International	The United States
QPR ProcessAnalyzer	QPR software	Finland
Aris Business Process Analysis	Software AG	Germany
Fujitsu Process Analytics	Fujitsu	Japan
XMAnalyzer	XMPro	The United States
StereoLOGIC Discovery Analyst	StereoLOGIC	The United States

ProM is a free open-source framework developed at the University of Eindhoven, where process mining was invented. It is a powerful process mining tool with lots of algorithms and features. It provides plug-ins for many different mining algorithms, as well as analysis, conversion and export modules.

Disco is a commercial tool also developed in Eindhoven at Fluxicon, a spin-off of the University of Eindhoven. It has a much more intuitive interface and integrated import functionality for CSV or Excel files. Its algorithm is based on the Fuzzy Miner, also included in the ProM framework. ProM and Disco are stand-alone programs.

Celonis and Perceptive are online in-browser software, which can be an advantage in current network based environments.

QPR Process Analyzer can be opened in-browser or downloaded as a plug-in for MS Excel. It offers the possibility to load data directly from a database and one advantage of this tool is that it is possible to use all features of MS Excel (graphs, lay-out) on top of the process mining features.

ARIS PPM provides facilities such as quantitative measurement of objectives and visualisation of process instances. In addition an aggregate process view can be compiled from a variety of data sources and realised as a process graph.

Fujitsu was used for visualizing business process flows based on data collected from systems, is seamlessly integrated within Process Analytics Software. It provides a holistic, end-to-end view of existing operations and business processes to reveal critical process intersections and hidden problems such as bottlenecks, delays, repetitions and failures.

XMAnalyzer was released by XMPro with process-mining information as part of their iBOS

(Intelligent Business Operations Server) solution. It has the Ability to analyze the tangible sequence flow of processes based on transactions, events or activities versus predetermined workflows. Also, it is easily accessible to end-users from their daily work areas.

StereoLOGIC Discovery Analyst is the first Automated Business Process Discovery solution on the market which automatically extracts business processes from business applications in real time, and speeds the creation of high quality models with powerful visualization and comparison capabilities.

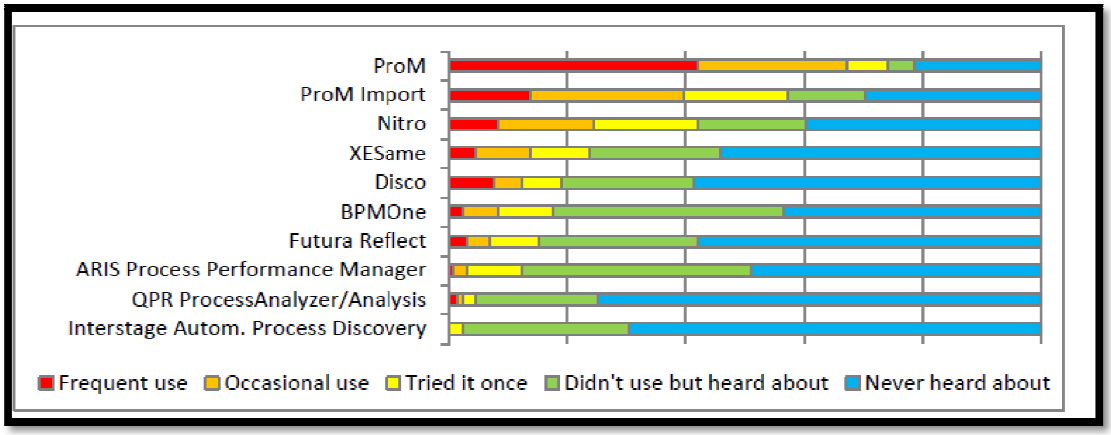

Figure 1.2: Tools for Process Mining

6. Benefits of process mining

1).The primary observes benefit of process mining is its objectivity.

2).The another benefit is its speed i.e., most process mining techniques get fast results after we have the right data in place.

3) It has the possibility to find exceptions and check conformance. Not only it helps in finding errors, but it also helps in identifying the causes for certain deviations.

4).It has the possibility to take different views on the same process or the data is also much appreciated.

7. Associated threat

There are important threats that need to be addressed; these highlight that process mining is an emerging discipline. The main challenges that encounter are outlined by van der Aalast (2004a) as listed. This list is not expected to be complete and, over time, new challenges may emerge or existing challenges may disappear due to advance in process mining.

7.1. Finding, Merging, and Cleaning Event Data:

In order to extract event data suitable for process mining considerable efforts are still taken. There are, several hurdles which need to be overcome:

- Data may be distributed over a variety of sources. For this the information needs to be merged, which tends to be problematic when different identifiers are used in the different data sources.
- Event data are often "object centric" rather than "process centric".
- Event data may be incomplete.
- An event log may contain outliers, i.e., exceptional behaviour also referred to as noise.
- Logs may contain events at different levels of granularity.

7.2. Dealing with Complex Event Logs Having Diverse Characteristics

Event logs are having very different characteristics. Some event logs may be extremely large making it difficult to handle them whereas; other event logs are so small that not enough data is available to make reliable conclusions. Therefore, additional efforts are needed to improve performance and scalability.

Also event logs contain only sample behaviour, they should not be assumed to be complete. Therefore, it will be challenging to deal with small event logs with a lot of variability.

7.3. Creating Representative Benchmarks

Process mining is an emerging technology. This explains why good benchmarks are still missing. For classical data mining techniques, many good benchmarks are available. These benchmarks have stimulated tool providers and researchers to improve the performance of their techniques.

On the one hand, there should be benchmarks based on real-life data sets. On the other hand, there is the need to create synthetic datasets capturing particular characteristics. Such synthetic datasets help to develop process mining techniques that are tailored towards incomplete event logs, noisy event logs, or specific populations of processes.

7.4. Dealing with Concept Drift

The term concept drift refers to the situation in which the process is changing while being analyzed. For instance, in the beginning of the event log two activities may be concurrent whereas later in the log these activities become sequential. Processes may change due to periodic/seasonal changes (e.g., "in December there is more demand" or "on Friday afternoon there are fewer employees available") or due to changing conditions (e.g., "the market is getting more competitive"). Such changes impact processes and it is vital to detect and analyze them.

Concept drift in a process can be discovered by splitting the event log into smaller logs and analyzing the "footprints" of the smaller logs. Such "second order" analysis requires much more event data.

7.5. Improving the Representational Bias Used for Process Discovery

A process discovery technique produces a model using a particular language (e.g., BPMN or Petri nets). However, it is important to separate the visualization of the result from the representation used during the actual discovery process.

The selection of a target language often encompasses several implicit assumptions. It limits the search space; processes that cannot be represented by the target language cannot be discovered. This so-called "representational bias" used during the discovery process should be a conscious choice and should not be (only) driven by the preferred graphical representation.

7.6. Balancing Between Quality Criteria such as Fitness, Simplicity, Precision, and Generalization Event logs are often far from being completed, i.e., only example behaviour is given. Process models typically allow for an exponential or even infinite number of different traces (in case of loops).

There are four competing quality dimensions: (a) fitness, (b) simplicity, (c) precision, and (d) generalization. A model with good fitness allows for most of the behaviour seen in the event log. A model has a perfect fitness if all traces in the log can be replayed by the model from beginning to end.

7.7. Cross-Organizational Mining

Traditionally, process mining is applied within a single organization. However, as service technology, supply-chain integration, and cloud computing becomes more widespread, there are scenarios where the event logs of multiple organizations are available for analysis.

In principle, there are two settings for cross-organizational process mining. First of all, we may consider the collaborative setting where different organizations work together to handle process instances. One can think of such a cross-organizational process as a "jigsaw puzzle", i.e., the overall process is cut into parts and distributed over organizations that need to cooperate to successfully complete cases.

Second, we may also consider the setting where different organizations are essentially executing the same process while sharing experiences, knowledge, or a common infrastructure.

7.8 Providing Operational Support

Today, many data sources are updated in (near) real-time and sufficient computing power is available to analyze events when they occur. Therefore, process mining should not be restricted to off-line analysis and can also be used for online operational support. Three operational support activities can be identified: detect, predict, and recommend. The moment a case deviates from the predefined process, this can be detected and the system can generate an alert. Often one would like to generate such notifications immediately (to still

be able to influence things) and not in an off-line fashion.

7.9 Combining Process Mining With Other Types of Analysis

Process mining techniques can be used to learn a simulation model based on historical data. Subsequently, the simulation model can be used to provide operational support. Because of the close connection between event log and model, the model can be used to replay history and one can start simulations from the current state thus providing a "fast forward button" into the future based on live data. Similarly, it is desirable to combine process mining with visual analytics. By combining automated process mining techniques with interactive visual analytics; it is possible to extract more insights from event data

7.10 Improving Usability for Non-Experts

One of the goals of process mining is to create "living process models", i.e., process models that are used on a daily basis rather than static models that end up in some archive. New event data can be used to discover emerging behaviour.

The link between event data and process models allows for the projection of the current state and recent activities onto up-to-date models. Hence, end-users can interact with the results of process interfaces.

7.11. Improving Understand ability for NonExperts

If it is easy to generate process mining results, it does not mean that the results are actually useful. The user may have problems understanding the output or is tempted to infer incorrect conclusions. To avoid such problems, the results should be presented using a suitable representation Moreover; the trust worthiness of the results should always be clearly indicated. There may be too little data to justify particular conclusions.

8. Disadvantages of process mining

- A major issue with process mining seems to be the search for data. It is difficult to find the right data of right quality and to fit it in the right structure.
- The lack of documentation and intuitiveness is also a drawback of current techniques.
- Most process mining techniques are very complex.

9. Application Areas

9.1 Process Discovery and Enhancement

A major area of application for process mining is the **discovery** of formerly unknown process models for the purpose of **analysis or optimization** (Van der Alast et al. 2012). Practitioners have since primarily focused on designing and implementing processes and getting them to work.

For traditional process modelling necessary information is primarily gathered by interviewing, workshops or similar manual techniques that require the interaction of persons. This leaves room for interpretation and the tendency that ideal models are created based on often overly optimistic assumptions.

9.2 Conformance Checking

A specific type of analysis in process mining is conformance checking (Adriansyah et al. 2011). The assumption for being able to conduct conformance checking is the existence of a process model that represents the desired process. For this purpose it does not matter how the model was generated either by traditional modelling or by process mining.

In general **local diagnostics** can be calculated that highlight the nodes in the model where deviations took place and **global conformance measures** that quantify the overall conformance of the model and the event log.

9.3 Compliance Checking

Compliance refers to the adherence of internal or external rules.. Compliance checking deals with investigating if relevant rules are followed. It is important in the context of internal or external audits. Process mining offers new and rigorous possibilities for compliance checking (Ramezani et al. 2012).

Compliance checking is a relatively novel field of research in the context of process mining. It can be distinguished as:

1. **Pre-runtime compliance** checking
2. **Runtime compliance** checking
3. **Post-runtime compliance** checking

Pre-runtime compliance checking is conducted when processes are designed or redesigned and implemented. Runtime compliance checks if violations occur when a business transaction is processed.

Post-runtime compliance is applied over a certain period of time, when the transactions have already taken place.

9.4 Organizational Mining

The aim of Organizational mining is to analyze information that is relevant from an organizational perspective. It involves the discovery of social networks, organizational structures and resource behaviour (Song and Van der Aalst 2008). Importance, distance, or centrality of individual resources type metrics can be computed.

References

1. Van der Aalst, W., Weijters, A. and Maruster, L. (2004) "Workflow mining: discovering process models from event logs", IEEE Transactions on Knowledge and Data Engineering, 16(9),1128–1142.
2. Rozinat, A. and van der Aalst, W. (2006a) "Conformance testing: measuring the fit and appropriateness of event logs and process models", BPM 2005 Workshops (Workshop on Business Process Intelligence), 3812, 163–176.
3. Rozinat, A. and van der Aalst, W. (2006b) "Decision mining in ProM", International Conference on Business Process Management (BPM 2006), 4102, 420–425.
4. Song, M., and Van der Aalst, W. M. P. (2008). "Towards comprehensive support for organizational mining," Decision Support Systems 46(1), 300–317.
5. Van der Aalst W.M.P., Adriansyan,A.,Karaia,A.,de Medeiros,A.K.A.,Arciernt,F., et al. (2012)"Process Mining Manifesto.Proc.BPM"11 Workshops,LNBIP 99,169-194 Springer.
6. Ramezani, E., Fahland, D., and Aalst, W. M. P. van der.(2012). "Where Did I Misbehave? Diagnostic Information in Compliance Checking," In Business Process Management Lecture Notes in Computer Science, A. Barros, A. Gal, and E. Kindler (eds.), 7481 Springer, 262–278.
7. Adriansyah, A., Van Dongen, B. F., and Van der Aalst, W. M. P. (2011). "Conformance checking using cost-based fitness analysis," In Enterprise Distributed Object Computing Conference (EDOC), 15th IEEE International, 55–64.
8. Gupta, E. (2014),"Process Mining Algorithms" International Journal of Advance Research in Science and Engineering 3(11), .401-412.
9. Dunham. M. H., (2002), "Data Mining: Introductory and Advanced Topics", Prentice-Hall.
10. Tiwari A. and Turner J. C.,(2008), "A review of business process mining: state-of-the-art and future trends", Process Management Journal,14 (1), 5-22.