Volume 5, Issue 2, February 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey Paper on Process Mining

Vidya N. Chavada, Pratik Kumar
CSE Department & GTU
Gujarat, India

Abstract— Process mining is originated from the fact that the modern information systems systematically record and maintain history of the process which they monitor and support. Business intelligence aims to support and improve decision making processes by providing methods and tools for analyzing data. Process mining builds the bridge between data mining as a business intelligence approach and business process management. Its primary objective is the discovery of process models based on available event log data. The discovered process models can be used for a variety of analysis purposes. In this paper challenges, different process mining algorithms, classification of process mining techniques are explained.

Keywords— Process Mining, Conformance Checking, Compliance Checking, Concept Drift, Event Log.

I. INTRODUCTION

Process Mining act as a bridge between data mining and business process modelling [6]. It is also known as process management technique based on event logs that allows for the analysis of business processes.Its main aim is discovering, monitoring and improving real processes by extracting knowledge from the event logs recorded by the various information systems. By extracting knowledge from event logs Process mining encompasses "techniques, tools and methods to discover, monitor and improve real processes [3]. Business processes in information systems is used for the data that is generated during the execution of reconstructing process models. These models are useful for analyzing and optimizing processes. Its starts from the event log, which is the collection of events. We can say that event can be related to process instances also called cases. The all events in the one case are ordered. Therefore, process instances are often represented as a trace over the set of activities. [1] In addition, events can have attributes such as timestamps, associated resources, transactional information and data attributes.

In the following section described Process Mining and its background. It consists of some different techniques, classifications and challenges in process mining.

II. BACKGROUND

Difference between process mining and data mining are: There are varieties of data mining methods have been developed (Classification, Regression, Clustering and Pattern discovery etc.). Process mining has a set of distinctive techniques, methods that produces different end results. Process mining does the process centric examination of the data. Data mining fails to offer the process centric analysis of the data. Data mining techniques make few assumptions about the arrangement of input data than process mining techniques. Process models exhibiting parallelism are incomparable to simple data mining structures such as decision trees and association rule [1].

Data Mining derives its name from the similarities between searching for valuable business information in a large database [9].Data mining involves six common classes of tasks known as anomaly detection, Association Rule Learning, Clustering, Classification, Regression etc. The aim of process mining is the construction of process models based on available logging data.

Fig. provides an overview of the different process mining activities. Before being able to apply any process mining technique it is necessary to have access to the data. It needs to be extracted from the relevant information systems. Data entries might need to be composed in a meaningful manner for the extraction. Another obstacle is the amount of data [2]. Before the extracted event log can be used it needs to be filtered and loaded into the process mining software. There are different reasons why filtering is necessary. Information systems are not free of errors. Errors can result from malfunctioning programs but also from user disruption or hardware failures that leads to erroneous records in the event log. Filtering is necessary to curtail the event log so that it only contains events that belong to the process. Such a filtering needs to be conducted carefully because it can lead to truncated process instances as well. Data filtering and loading is commonly supported by software tools and performed in a single step.

The mining includes the discovery of relationships in the event log whereas the reconstruction produces a process model as a graphical representation. The mining and reconstruction are commonly provided by the same software tool in a single step. When the process models are mined and reconstructed they can be used for the intended purpose. We summarize this step with the term analysis.

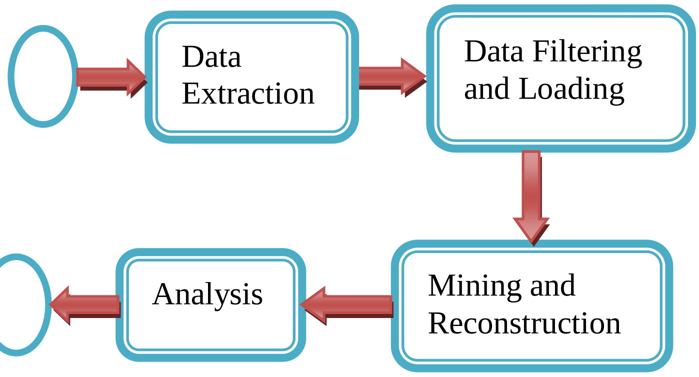

Fig. 1 Process mining activities

III. CHALLENGES

A. Noise and Incompleteness

Erroneous records in the event log should be distinguished from a phenomenon called noise which leads to increased complexity in the process model. Process mining approaches should therefore be able to handle or filter out noise [2].

B. Dealing with Complex Event Logs Having Diverse Characteristics

Events in an event log are already mapped to cases. Supporting and automating the execution of business processes and therefore most likely also store high quality event data that can easily be used for process mining [2]. Thus the quality of an event log depends on the source system's ability to record process relevant data.

C. Finding, Merging, and Cleaning Event Data

Data may be distributed over a variety of sources. This information needs to be merged. This 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. A common problem is that events do not explicitly point to process instances. An event log may contain outliers,i.e., exceptional behavior also referred to as noise. Logs may contain events at different levels of granularity. Events occur in a particular context (weather, workload, day of the week, etc.). This context may explain certain phenomena, e.g., the response time is longer than usual because of work-in-progress or holidays.

D. 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. Such changes impact processes and it is vital to detect and analyze them[2].

E. Balancing Between Quality Criteria such as Fitness, Simplicity, Precision and Generalization

Fitness addresses the ability of a model to replay all behavior recorded in the event log. Simplicity means that the simplest model that can explain the observed behavior should be preferred. Precision requires that the model does not allow additional behavior very different from the behavior which is recorded in the event log. Generalization means that a process model is not exclusively restricted to display the eventually limited record of observed behavior in the event log but that it pro vides an abstraction and generalizes from individual process instances[2].

F. 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 consider the collaborative setting where different organizations work together to handle process instances. Second consider the setting where different organizations are essentially executing the same process while sharing experiences, Knowledge or a common infrastructure.

G. Providing Operational Support

Initially, the focus of process mining was on the analysis of historical data. Today, however, many data source 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.

H. Combining Process Mining With Other Types of Analysis

Operations management, and in particular operations research, is a branch of management science heavily relying on modeling. Here a variety of mathematical models ranging from linear programming and project planning to queuing models, Markov chains, and simulation are used.

IV. CLASSIFICATION

There are three classes of process mining techniques.

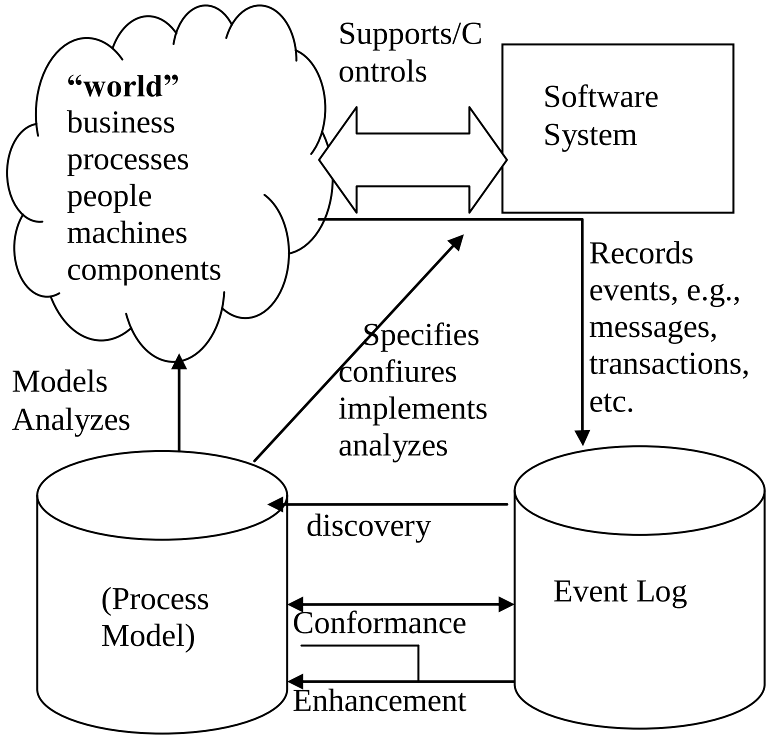

Fig 2. Types of Process mining

A. Discovery

There is no a priori model. A major area of application for process mining is the discovery of formerly unknown process models for the purpose of analysis or optimization.[1] Based on an event log some model is constructed a process model can be discovered based on low-level events. There are many techniques to automatically construct process models based some event log.

B. Conformance analysis

There is an a priori model. Event logs can be replayed to identify conform or deviant behavior it is a specific type of analysis in process mining. Event log is compared with the process model for identifying conform or deviant behavior. When conducting conformance checking it should be kept in mind that not every deviation needs to be negative and should therefore be eliminated. Major deviations from the ideal model might also mean that the model itself does not reflect real world circumstances and requirements.

C. Enhancement

There is a prior model also. This model is extended with a new aspect or perspective, i.e., the goal is not to check conformance but to enrich the model.

V. PROCESS MINING ALGORITHM

The main component in process mining is the mining algorithm. It determines how the process models are created. A broad variety of mining algorithms does exist. The following three categories will be discussed in more detail.

A. Heuristic mining algorithms

Determinism means that an algorithm only produces defined and reproducible results. It always delivers the same result for the same input [4]. Heuristic mining also uses deterministic algorithms but they incorporate frequencies of events and traces for reconstructing a process model [9].

B. Fuzzy Miner

Fuzzy Miner mines behaviour of less structured processes.[11] It applies a variety of techniques, such as removing unimportant edges, clustering highly correlated nodes in to a single node, and removing isolated node clusters.

C. Genetic mining algorithms

Genetics Miner uses a genetic algorithm to mine a Petri net representation of the process model from execution traces. The algorithm employs a search technique that mimics the evolution of biological systems. Although 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[11].

Table 1 Comparative Study of Mining Algorithm [9]

	Heuristic Miner	**Fuzzy Miner**	**Genetic Miner**
Description	Provides a view of scientific workflows	Provides a zoom able view of scientific workflows	Provides a view of frequency for both tasks and succession
Strategy	Work based on local strategy technique to build a model	Work based on local and global strategy technique to build a model	Work based on global strategy technique to build a model
Output	Heuristic net	Fuzzy model	Petrinent graph
When to use it	When you have real-life data with not too many different events	When you have complex and unstructured data	When you need to generate a random population of process model
Behavior	Take frequencies into account	Uses dependency graph representation	Uses mimic natural evolutions
Format	HM can mine Unstructured process	FM can mine less structured process	Gm can mine both structured and unstructured process
Log Data	Can handle incomplete logs to a certain extent.	Can handle incomplete logs.	Can easily handle incomplete logs.

VI. RELATED WORK

Over the last two decades many researchers have been

Working on process flexibility, e.g., making workflow systems adaptive. Ploesser et al. [11] have classified business process changes into three broad categories: 1) sudden; 2) anticipatory; and 3) evolutionary. Despite the many publications on flexibility, most process mining techniques assume a process to be in a steady state. This approach uses process mining to provide an aggregated overview of all changes that have happened so far. This approach, however, assumes that change logs are available, i.e., modifications of the workflow model are recorded.

Concept drift has been studied in both supervised and unsupervised settings and has been shown to be important in many applications. The first ever published work on concept drift in process mining was by JC Bose et al.[] , which deals with handling (detection, localization, characterization) the sudden drift in control flow perspective of the process in offline manner. Concept drift focuses on changes in simple structures such as variables, concept drift in process mining deals with changes to complex artifacts such as process models describing concurrency, choices, loops. Although experiences from data mining and machine learning can be used to investigate concept drift in process mining, the complexity of process models and the nature of process change pose new challenges [7].

We can differentiate between two broad classes of dealing with concept drifts when analyzing event logs.

1) Offline analysis: This refers to the scenario where the

Presence of changes or the occurrence of drifts need not be uncovered in a real time. This method considers only recoded data (historical data). This is appropriate in cases where the detection of changes is mostly used in post-mortem analysis, the results of which can be considered when designing/improving processes for later deployment.

2) Online analysis: This refers to the scenario where changes need to be discovered in near real time. It consider recorded data, completed process traces while analysis, pre mortem data. This is appropriate in cases where an organization would be more interested in knowing a change in the behaviour of their customers or a change in demand as and when it is happening. Such real-time triggers (alarms) will enable organizations to take quick remedial actions and avoid any repercussions.

VII. CONCLUSION

Process mining builds the bridge between data mining and business process management. The increasing integration of information systems for supporting and the quality of an event log depends on the source system's ability to record process relevant Data Complexity. The data that is stored in the information systems can be used to mine and reconstruct business process models. The importance of process mining increases with the growing integration of various information systems. Process mining is still a young research discipline. The suture scope of this will be to have runtime changes in the process model with the help of these algorithms with monitoring, verification of requirements, and compliance enforcement.

REFERENCES

[1] Manoj kumar M V,Likewin Thomas, Annapa B, "Phenomenon of Concept Drift from Process Mining Insight"*Department of CSE,NITK*,978-1-4799-2572-8/14/$31.00 2014 IEEE.

[2] StB Prof. Dr. Nick Gehrke Nordakademie “Process mining” *Chair for Information Systems Köllner Chaussee 11,Michael Werner, Dipl.-Wirt.-Inf. University of Hamburg Chair for Information Systems , Max-Brauer-Allee* 60 D-22765 Hamburg.

[3] r. p. jagadeesh chandra bose, wil m. p. van der aalst, indr˙e žliobait˙e, and mykola pechenizkiy, “Dealing With Concept Drifts In Process Mining”**,** *ieee transactions on neural networks and learning systems, vol. 25, no. 1,* january 2014.

[4] Jianmin Wang, Raymond K. Wong, Jianwei Ding, Qinlong Guo, And Lijie Wen,“Efficient Selection Of Process Mining Algorithms**”**, *Ieee Transactions On Services Computing, Vol. 6, No. 4,* October-December 2013.

[5] R.P. Jagadeesh Chandra Bose, Wil M.P. van der Aalst, Indr_Zliobait, and Mykola Pechenizkiy, “Handling Concept Drift in Process Mining”, *Department of Mathematics and Computer Science, University of Technology, The Netherlands*.

[6] Van der Aalst, W. M. P. “Process Mining: Discovery, Conformance and Enhancement of Business Processe*”*, (*1st Edition) Heidelberg: Springer*.

[7] R. P. J. C. Bose, W. M. P. van der Aalst, I. Žliobait˙e, and M. Pechenizkiy,“Handling concept drift in process mining,” in *Proc. Int. CAiSE*, 2011,pp. 391–405

[8] J. Carmona and R. Gavaldà, “Online techniques for dealing with concept drift in process mining,” in *Proc. Int. Conf. IDA*, 2012, pp. 90–102

[9] Asst. Prof. Esmita .P. Gupta, “Process Mining A Comparative Study” *International Journal of Advanced Research in Computer and Communication EngineeringVol. 3, Issue 11,* November 2014.

[10] K. Ploesser, J. C. Recker, and M. Rosemann, “Towards a classification and lifecycle of business process change,” in *Proc. BPMDS*, vol. 8. 2008, pp. 1–9.

[11] Geetika T. Lakshmanan and Rania Khalaf, “Levaragig process mining techniques”, *IBM T.J. Watson Research Center,* IT Pro September/October 2013

