IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Process Mining - A Comprehensive Review

S. Vijayarani[1], A. Sakila[2], R. Ramya[3]

[1]Assistant Professor, Department of Computer Science, Bharathiar University, Coimbatore, India
[2]Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore, India
[3]PG Student, Department of Computer Science, Bharathiar University, Coimbatore, India

Available online at: www.ijcseonline.org

Accepted: 19/Jul/2018, Published: 31/Jul/2018

***Abstract* –** Process mining is an emerging research field of computer science, which utilizes automatically generated event logs in information system. An event log contains time stamped record for events which refer to activities under taken for specific cases. Event log also store additional information regarding event such as resources linked to the activity, or data elements recorded with the event. To allow these events to happen in parallel to the current state of a process may be represented by multiple nodes. Those process model can follow the notation of petri nets and Business process Modelling Notation (BPMN), widely used for their improvement. Process model guidance is an important model feature by which the software process is orchestrated. Process mining techniques have been used in business process domain with no focus on the software engineering processes. These techniques have produced faster results and the ability to check conformance and compliance. This paper discussed the basic concepts, applications, frameworks, types, algorithms and research issues of process mining.

***Keywords* —** Process Mining(PM), Educaional Process Mining(EPM), Process Mining techniques, Business Process Modelling(BPM), Event Logs.

I. INTRODUCTION

Process mining techniques are used for extracting information from event logs. It is used in the field of process management that supports the analysis of business process. It is closely related to Business Activity Monitoring (BAM), Business Process Management (BPM) and Business Process Intelligence (BPI) [1]. Process mining aims to bridges the gap between traditional BPM and BPI. An input for all process mining is event data which records the information about the execution of business process. Process mining aims to discover, monitor and improve real process by extracting knowledge from event logs which are readily available in today's information system. A spectacular growth of event data and process mining technique have matured significantly [3]. Management trend related to process improvement and compliance can now benefit from process mining. Starting point of process mining is an "event log". Each event in such a log refers to an "activity" and is related to particular "process instance". The events belonging to a case are ordered and can be seen as one run of the process. Event logs may store additional information about events.

Process mining generates number of scientific and practical challenges. The goal of process mining is to extract an explicit process model from event logs, i.e. the challenge is to create a process model, given a log with events such as the model is consistent with the observed dynamic behaviour. Process mining also focuses on finding causal relationship between activities [2]. Business Process Modelling (BPM) used hand-made models whereas process mining is based on facts. Based on observed behaviour recorded in event logs, intelligent techniques are used to extract knowledge. Process mining is process-centric, truly intelligent and fact based. Process mining is a process management technique, analyses the business processes based on event logs. The basic idea is to extract knowledge from event logs recorded by an information system. This knowledge is delivered in the form of business process models. Process mining algorithms are used to mine business process models using process logs [4]. The mined models will then be compared against the original process model of the enterprise of conformance checking.

The remaining portion of the paper is organized as follows. Section 2 presents the application of process mining. Section 3 represents the framework of process mining. Process mining techniques gives in section 4. Section 5 provides the activities in process mining. Research challenges and issues in process mining are discussed in Section 6. Section 7 gives the conclusion.

II. RELATED WORK

Gabriel M.Veiga, et al. [19] compared the existing business process mining tools. That provides a graphical representation of an each process, though the method and notation used in each representation may differ. The completeness of event logs and the amount of errors that contains the challenges to the process mining tool. The practice of business process mining attempts to reconstruct the complete process models from data logs, it containing real process execution data.

W.M.P.Van der aalst, et al. [17] analysed the event logs generated by workflow management system of the organization involved. Most of the workflow is focusing on modelling, verification, simulation, and enactment rather than process mining. The issue of process mining in the context of workflow management using an inductive approach. Business Process Intelligence (BPI) and Business Activity Monitoring (BAM) are used in Business process mining. It aims at the automatic construction of models explaining the behaviour observed in the event logs.

Awatef Hicheur Cairns, et.al [12] proposed a process mining technique for discovering interaction patterns from email datasets. In the discovered patterns from the data set, students in their collaborative learning tasks, communication actions and online discussions. Their proposed technique may be tested on a distributed environment connected to several data sources and applications. To enhance the usability, they also working on designing an intuitive graphical interface for automatically set parameters and suggests suitable types of analysis.

III. APPLICATIONS OF PROCESS MINING

The goal of process mining is to extract information from event logs. During process mining, specialized data mining algorithms are applied to event log data in order to identify trends, patterns and details contained in the event log recorded by an information system. Process mining aims to improve process efficiency and understanding of processes. It describes a family of a posteriori analysis techniques exploiting the information recorded in the event logs. Process mining can be used in the context of Business Process Intelligence (BPI). BPI set includes "BPI Process Mining Engine". This does not provide any technique, it uses generic mining tools such as SAS Enterprise Miner for the generation of decision trees. In order to do workflow mining it is convenient to have a so-called "Process Data Warehouse" [5].

Process mining applied to education data is called Educational Process Mining (EPM). It is the adaptation framework of the generic framework of the process mining. The components are:

i) Educational world.
ii) Virtual learning environment.
iii) Event logs.

Educational world: Two participants play an important role in any e-learning activity; teachers and students. Teacher supplies appropriate resources to ensure student success. Students are the essential parts of any e-learning activity, interacting with other participants and with the system itself. It aims to make unexpressed knowledge explicitly and to facilitate better understanding of the educational process

Virtual Learning Environment (VLE): It supplies basic structure and resources where the participants learning action and interaction occur. It also logs the event that occur during e-learning process. It provides teachers or researchers with basic tools for analysing students learning based on marks evaluation, number of activities done, forum participation and last log in. Sometimes it does not provide specific tools because that would allow educators to thoroughly assess the overall student learning process. A Virtual Learning Environment (VLE) in educational technology is a web-based platform for the digital aspect to the course of study.

Components of VLE

- Content management- Creation, Storage, Access to and use of learning resources.
- Curriculum mapping and planning- Lesson planning, assessment and personalisation of the learning experiance.
- Learner engagement and administration- Managed access to learner information,resources and tracking of progress and achievement.
- Communication and collaboration- Emails, chats, notices, wikis, blogs, etc..
- Real time communication- Live video conferencing and audio conferencing.

There are three major applications used in Process Mining. They are:

i) Process mining used during walkthroughs.
ii) Process mining used for sampling.
iii) Process Mmining used for compilance checking.

If a process model does not exist the software will perform business process discovery to create the model automatically, by using Artificial Intelligence or Machine Learning. The technology is often applied to the most common and complex business processes executed in organisations. An organisation might use process mining software to find the cause of unexpected delays in invoice processing [6].

IV. FRAMEWORK AND CONCEPTS

Process mining aims to extract knowledge from "process execution logs". The traditional process mining performs the analysis and design processes by using business process modelling tool. Figure 1 represents the framework for process mining.

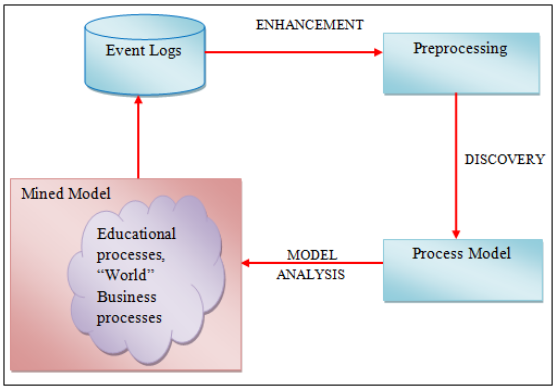

Figure 1 Framework for Process Mining

Pre-processing: It is a program that process its input data to produce output that is used as input to another program like a compiler. It is used in machine learning and data mining to make input data easier to work. There are five different tasks included in data pre-processing. They are:

- **Data Cleaning:** It fills in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
- **Data integration:** It is using multiple databases, data cubes, or files.
- **Data Transformation:** It includes normalization and aggregation.
- **Data Reduction:**It reduce the volume but producing the same or similar analytical results.
- **Data Discretization:** It is a part of data reduction, replacing numerical attributes with nominal ones.

Event logs: Process mining assumes the existence of an event log where each event refers to a case, an activity and a point in time. An event log can be seen as a collection of cases and it can be seen as trace/ sequence of events. These are files that record events that occur in VLE and are normally stored in database. They contain a large amount of data about the educational agent's interaction with the VLE. Event logs need to be transformed into a particular file format in order to be used by a specific PM tool. It records events taking place in the execution of the system in order to provide an audit trail that can be used to understand the activities of complex system. It can also be useful to combine log file entries from multiple sources [9].

Process mining techniques: It is defined as a technique to discover, monitor and improve real process by extracting knowledge from event logs readily available in information systems. Events may be stored in database tables, message logs, mail archives, transaction logs, and other data sources. Processing mining techniques are described in the section IV.

V. PROCESS MINING TECHNIQUES

There are many process mining techniques to meet process mining with different requirement. Popular process mining algorithms are alpha-miner, alpha++-miner and heuristic miner. Process discovery techniques build a process on an event log by capturing the behaviour seen in the log. They focus on the control-flow perspective of process. There are many algorithms in process mining for discovering underlying processes from event logs. Some of the process mining algorithm are described below.

Alpha Algorithm: It aimed at reconstructing a set of sequence of events. It is able to deal with concurrency. It takes an event log as input and calculates the ordering relation of the event contained in the log. It constructs workflow nets with special properties from event logs. Each transition in the net corresponds to an observed tasks.

Heuristic Miner Algorithm: It focuses on the control flow perspective and creates a model in Heuristic nets. It is also used in deterministic algorithm but it extends alpha algorithm by considering the frequency trace in the event log.

Steps of Heuristic Miner Algorithm

1. Read a log.
2. Get the set of tasks.
3. Order relation based on relation.
4. Build the net based on infrared relation.
5. Output the net.

Genetic Miner Algorithm: It uses an evolutionary approach that uses the process of natural evolution. The algorithm can mine the process model that contain all common structural constructs and able to handle noise. It provides process models built on casual matrixes. The algorithm starts by randomly distributing the finite number of points into this search space. Every point in the search space is called an individual and the finite set of point at a moment in time is called Population. It tackles problems such as noise, incomplete data, non-free choice constructs, hidden activities, concurrency and duplicate activities [7].

Fuzzy Miner Algorithm: It is one of the newer process discovery algorithm. Its process is to empower users to interactively explore processes from event logs. Fuzzy miner is suitable for mining less-structured process which exhibit a large amount of unstructured and conflicting behaviour. The boundary of application can vary considerably according to context or conditions, instead of being fixed once and for all. The new conceptual method make it possible to identify, to measure and respond to fine graduations of significance, with great precision. The new conceptual method make it possible to identify, to measure and the significance, with great precision. It is the first algorithm which directly address the problems of large number of activities and highly unstructured behaviour [12].

VI. ACTIVITIES OF PROCESS MINING

Process mining starts by gathering information about the processes. It is possible to record events such that the information about the order in which the events of the case are executed. The information system using transactional system such as ERP, CRM, B2B, SCM and WFM system offer information in some forms. This process log is used to construct a specification, which adequately models the behaviour registered. Figure2 shows the activities of process mining.

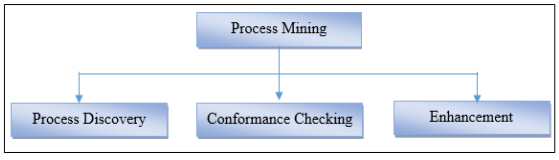

Figure 2 Activities of process mining

Process mining technique can be categorized under these three activities, they are described below.

1. **Process discovery:** Process mining has been focusing on discovery. To construct a comprehensive process model, it can be able to reproduce the behavior seen in the log file.In the event log, based on the low-level events, a new model is constructed or discovered. Many techniques are available for automatic construction of process model by using event logs. It takes no-apriori process schema, and some schema is constrcuted based on the particular event log [13].

2. **Conformance checking:** This is an a-priori model. This model is used to check the reality conforms to the model.Conformance checking may be used to detect deviations, to locate these deviations and to measure the severity of deviation. The existing model is compared with the process event logs and to find the discrepancies between the analyzed log and model. Some priori model is used to project the potential bottlenecks.It is used to detect deviation to enrich the model. It finds the deviation between observed behaviors in event log and the generated process model. It takes an existing process schema. It is used as a reference and to check if reality conforms to the schema [14].

3. **Enhancement:** This model is extended with a new aspect or perspective and aims to improve or extend a given process model based on information extracted from a specific event log. It takes the existing process schema. It is extended with a new aspect or perspective especially the goal is not to check conformance but tries to extract new information from it [15].

VII. CHALLENGES AND ISSUES IN PROCESS MINING

Process Mining is an important technology for modern organisation that need to manage non-trivial operational process. There is an incredible growth of event data, process and information need are to be aligned perfectly in order to meet requirements related to compliance, efficiency and customer service. Some of the most important process mining challenges are:

Finding, Merging and Cleaning event data: When extracting event data it must be suitable for process mining;

here several challenges need to be addressed. Data may be distributed over a variety of sources and ensure that it may not be incomplete.

Dealing with Complex Event logs: Event logs may have very different characteristics. Some event logs may be extremely large, making them difficult to handle event logs are so small not enough data is available to make reliable conclusion. If the event logs are extremely large, then it is very difficult to handle. If it is small then it gives reliable conclusion.

Creating representing benchmarks: Good benchmark data sets and representative quality criteria are needed to compare and improve the various tools and algorithms.

Dealing with concept drifts: The process may be changing while being analysed. Understanding such concept drifts is of prime importance for the management of process.

Cross-organisational mining: There are various use cases where event logs are available for analysis in the multiple organization. Some organisations work together to handle process instances or organisations. Traditional process mining techniques typically consider one event log in one organisation.

Providing operational support: Process mining is not restricted to off-line analysis and can be used for online operational support. There are three additional support: Delete, Predict and Recommend.

Improving usability for non-experts: The challenge is to hide the sophisticated process mining algorithm behind user-friendly interfaces that automatically set parameters and suggest suitable types of analysis.

Combining process mining with other: The challenge is to combine automated process mining technique with other analysis approaches to extract more insights from event data.

Improving usability for Non-Experts: The challenge is to hide the sophisticated process mining algorithm behind user-friendly interfaces that automatically set parameters and suggest suitable types of analysis.

Improving understand ability for Non-experts: the user may have problems to understand the output to infer incorrect conclusion. To avoid such problems, the result should be presented using a suitable representation and the trustworthiness of the result should always be clearly indicated.

Disadvantages of process mining: A major issue with process mining seems to be 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 [18].

VIII CONCLUSION

Process mining is a new technology which enable evidence-based process analysis. This paper presents a comprehensive introduction of process mining concepts, applications, frameworks, and techniques. Process mining techniques are helpful for extracting information from event logs. Different types of activities and open research challenges and issues in process mining also discussed in this paper.

REFERENCES

[1]. AALST, W. VAN DER, "Process Mining: Discovery", Conformance and Enhancement of Business Processes.Springer-Verlag, Berlin, 2011.

[2]. AALST, W. VAN DER, SCHONENBERG, M and SONG, "Time Prediction Based on Process Mining.Information Systems"2011.

[3]. AALST, W. VAN DER and STAHL, "Modeling Business Processes: A Petri Net Oriented Approach", Business Process Management: Concepts, Languages, Architectures. Springer-Verlag,Berlin, 2011.

[4]. Wil van der Aalst chistian jan recker and Manfred Reichert, Eindhoven university of technology.

[5]. van der Aalst, W.M.P., Weijters, A.J.M.M., Maruster, L.: Workflow Mining: Dis-covering Process Models from Event Logs. IEEE Transactions on Knowledge andData Engineering 16 (2004) 11281142.

[6]. van der Aalst, W.M.P., van Hee, K.M.: "Workflow Management: Models, Methods,and Systems", MIT Press, Cambridge, Massachusets (2002)

[7]. Khodabandelou G, Hug C, Deneckere R, Salinesi C, "Processmining versus intention mining" Enterprise,Business-Process and Information Systems Modeling.Springer Berlin Heidelberg,Valencia, Spain; 2013,466–480.

[8]. Agrawal R, Srikant R. "Mining sequential patterns",Data Engineering Proceedings of the Eleventh International Conference. IEEE, Taipei, Taiwan; 1995, 3–14.

[9]. Nesbit JC, Zhou M, Xu Y, Winne P. "Advancing loganalysis of student interactions with cognitive tools",12th Biennial Conference of the European Associationfor Research on Learning and Insruction (EARLI). Budapest, Hungary; 2007, 2–20.

[10]. Reimann P, Frerejean J, Thompson K,"Using processmining to identify models of group decision making in chat data",9th international conference on Computer supported collaborative learning,Vol. 1. Rhodes, Greece; 2009, 98–107.
[11]. Washio T, Motoda H. State of the art of graph-baseddata mining. Acm Sigkdd Explor Newslett 2003,5:59–68.
[12]. Günther CW, Van Der Aalst WM, "Fuzzy mining–adaptive process simplification based on multiperspectivemetrics", International Conference onBusiness Process Management. Springer Berlin Heidelberg;Brisbane, Australia; 2007, 328–343.
[13]. Van der Aalst, W., Weijters, A. AndMaruster, L, "Workflow mining:discovering process models from eventlogs", IEEE Transactions on Knowledgeand Data Engineering, 16(9),1128–1142, 2004.
[14]. Rozinat, A. and van der Aalst, W. (2006a),"Conformance testing: measuring the fit andappropriateness of event logs and processmodels", Workshops (Workshopon Business Process Intelligence), 3812,163–176, 2005
[15]. Rozinat, A. and van der Aalst, W. (2006b)"Decision mining in ProM", InternationalConference on Business ProcessManagement (BPM 2006), 4102, 420–425.
[16]. Mekhala, "A review paper on process mining", International journal of engineering and techniques- volume 1 issue 4, july-aug 2015.
[17]. W.M.P.van der aalst and a.j.m.m.weijters, "Process Mining:A research Agenda", Eindhoven university of technology.
[18]. Van der aalst, W.M.P. 2011, "Process mining:Overview and oppurtunities", ACM trans. Management information system 99, article 99.
[19]. Gabriel M. Veiga and Diogo R. Ferreira, "Understanding Spaghetti Models with Sequence Clustering for ProM"

Authors Profile

Dr. S.Vijayarani is an Assistant Professor of Department of Computer Science at Bharathiar University, Coimbatore, India. She has obtained M.C.A., M.Phil., and Ph.D., in Computer Science. She has 10 years of teaching/research and 10 years of technical experience. Her research interests include data mining, privacy issues in data mining, text mining, web mining, data streams and information retrieval. She has published more than 80 research articles in national/international journals. She also presented research papers in international/national conferences. She has authored a book and guided more than 25 research scholars. She is a life member in professional bodies like CSI, ISCA, IAENG, IRED and UACEE.

Ms. A. Sakila is currently pursuing Ph.D in Computer Science at Bharathiar University, Coimbatore, India. She has Completed Master of Science in Computer Science and Master of Philosophy in Computer Science. Her research areas include data mining, image mining, process mining and big data. She has published research papers in international journals. She also presented research papers in national and international conferences. Her research areas include data mining, image mining.

Ms. R. Ramya is currently pursuing M.Sc Computer Science at Bharathiar University, Coimbatore, India. She has Completed Bachelor of Science in Computer Science. Her research areas include data mining, and process mining.

