Study On Spaghetti Process Mining with Concept Drift

N. Swapna[1*], L. Ramaparvathy[2]

[1]Dept. of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
[2]Dept. of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Abstract— Data science is the occupation of future, because organizations that are unable to use (big) data in a smart way will not survive. Process mining is a rising area that fills the gap between business process management techniques and data-centric analysis techniques such as machine learning and data mining. Process mining seeks the confrontation between event data and process models (hand-made or discovered automatically).It can be applied to any type of operational processes.Business process management is a top down approach. BPM starts by designing your process in high level model. Then you configure your system for managing and controlling the designed process. This system then coordinates work between employees and other resources in organization such that the organization is able to achieve the planned process. On the other end process mining analyzes process in a bottom-up fashion. That is we do not need to have model of process. Process mining uses the history data which is present in IT systems in the form of event data. Using this event data process mining generate process models as per the generated models organizations can take further steps to improve the models that are generated. As a result an organization cannot change the data but it can change the process in which the data is generated and hence work to meet the goals of organization. This paper gives an abstract view of process mining- algorithms used, applications, scope of process mining in diverse disciplines, research issues of Spaghetti process mining with concept drift.

Keywords— Business intelligence,business process management,operational processes, process mining, event data,process models,concept drift,Spaghetti process mining.

I. INTRODUCTION

Very huge amount of data is generated in just few minutes which was collected for few days until 2003. What kind of data is this? We are generating many different kinds of event data like- when we buy a cup of coffee with our credit card, when we make a phone call, etc. We are generating data because all kinds of things are being recorded. The challenge today is not to generate more data, but the challenge today is to turn this data into real value. We need to extract value from the data, apply process mining techniques[5] to it. Process mining focuses on the process the data is generated and hence gives needed result to the analyst. Hence process mining bridges the gap between business processes and big data by focussing on processes using real data.

In classical data mining people typically do not look at processes. Business intelligence is not sufficient tool (ex: anscombe data set) hence we use process mining. Process mining is the missing link between model based analyses, process model based analyses, and data oriented analyses like data mining .Its the glue between data and process. Its the glue between business people and IT people, business intelligence and business process management..

Where does Process mining stand?
As previously stated, business intelligence is defined as the methods and tools used by organizations to glean analytical findings from data. It also consists of how companies can gain information from big data and data mining. This means business intelligence is not confined to technology — it includes the business processes and data analysis procedures that facilitate the collection of big data.

Data mining falls under the umbrella term of "business intelligence," and can be considered a form of BI. Data mining can be considered a function of BI, used to collect relevant information and gain insights. Moreover, business intelligence could also be thought of as the result of data mining. As stated, business intelligence involves using data to acquire insights. Data mining business intelligence is the collection of necessary data, which will eventually lead to answers through in-depth analysis.

The link between data mining and business intelligence can be thought of as a cause-and-effect relationship. Data mining searches for the "what" (relevant data sets) and business intelligence processes uncover the "how" and "why" (insights). Analysts utilize data mining to find the information they need and use business intelligence to determine why it is important.

Process mining is a new and exciting field of this decade which combines business process management with data science. Using process mining techniques you can analyse and visualise business processes based on event data recorded in event logs. Process mining provides a critical, process-centric perspective on data, which is not

available with classical data mining or machine learning techniques. Spaghetti process mining is more interesting and important in view of process mining since using it we can analyze more complex event traces which cannot be done manually.
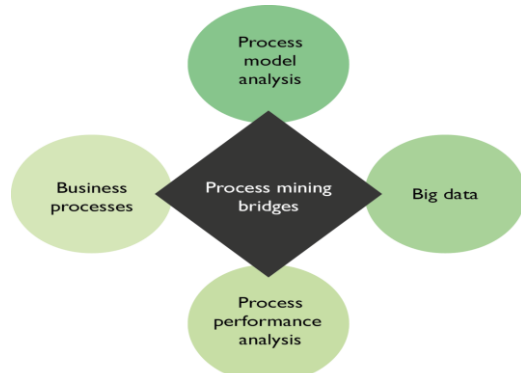

Fig.1

Process Intelligence

Process mining which fills the gap between process mining and business intelligence can now include Big Data analytics for business processes called process intelligence. PI can use the best practices of process mining and also use prediction with AI. Till now consultants are considered as best solution to business solution, but with PI it increases human interactions and continuously improves process execution.

The goal of this study is to provide an overview of process mining for both practitioners and process owners. The goal of the study is clear by focusing on below scenarios:
Where does process mining stand in relation to relate areas.
What does process mining exactly refer to?
Uses of process mining
Tools used for process mining.
Algorithms used for process mining.
Types of process mining(Lasanga and Spaghetti)
Overview of existing methods.

II. RELATED WORK

Martin Prodel et,al[1]in their work proposed an approach for process discovery from complex logs which combines various optimization techniques and process mining. They proposed tabu search algorithm which performs hierarchical classification on given process model and replays it compares the generated model with the one generated using Disco.[8]

Na Liu et.al[2] presented a mechanism to detect all four types of concept drifts – sudden drift, gradual drift, incremental drift and recurring drift. They proposed a frame work for online process concept drift detection from event streams. Their work involved in exploiting the relationship between activities in processes. They used three algorithms for detecting concept drift, first one to examines whether current model is stable or not. If not it uses algorithm2 for initialization. The second algorithm is

used to stabilize the current model. Algorithm 3 updates the current model.
Pros: Concept drifts are handled
Cons: Complexity of process is not taken into consideration.

Likewin Thomas et.al[3] proposed a framework to convert Spaghetti to Lasanga process. The proposed model is verified based on fitness value. A process that is simplified is considered as Lasanga when threshold value is 0.8 . Fitness value can be improved by decreasing the complexity of a given Sphagetti process.
Pros: Fitness value is considered
Cons:only small data set is taken

W.vander Aalst[4] shown a way of reducing complexity of Spaghetti process by applying fuzzy miner approach where Hierarchical clustering is applied to homogenous event logs.

Pros: Complexity of event log is reduced.
Cons: Event log in real time is not always homogenous .
Tugba Gurgen Erdogan and Ayca Tarhan [7] proposed a Goal driven evaluation system for healthcare Process.
Pros: Gives more insight into the process model.
Cons: Is more suitable for health care data sets.

III. METHODOLOGY

Uses of Process mining

Imagine a manager at a credit card application process in a bank. Customers can order credit cards at a local branch or online. When an employee verifies the application and risk manager conducts a credit check, if the card gets approved it is sent out. The problem is the customer start going to another bank because they are faster. 90% of the applications take more than 9 days. Why does it take so long? Process mining takes existing data from IT systems as starting point extracts different variations of the process and alternatively it turns them to understandable visualizations of your process. Because existing IT records are the places for process mining we can get visualization of the whole process. For example it shows remarkable deviation, real bottlenecks, and necessary rework. Using this customer can now get card in five days.

Process mining can be used in any field of study where we have event data related to the study. With process mining, you get a process model for these data. This way your real process and actual business rules can be discovered automatically. There are four software for process-mining tools namely [9] ProM, Disco, Celonis and My-Invenio.
ProM for RapidMiner: The proM framework and RapidMiner is now being used as a standard tool for process mining.

The recommended Process Mining Algorithms
1. Alpha algorithm: Alpha algorithm is the first process mining algorithm, it constructs a petrinet from the event log where each transition corresponds to an observed task.

Output:PetriNet

Pros: First proposed algorithm for process mining and hence satisfies basic requirements.
Cons: It cannot predict different type of transitions which are present in the workflow.

2. Heuristic miner: Heuristic Miner is the next to alpha algorithm, developed by Dr.Ton Weijters he used an heuristic approach to address various problems faced using alpha algorithm. It takes an event log as input and discovers model that describes control perspective of a process that was captured in event log showing the relationship between activities. The Heuristic miner finds relationships between activities which an alpha miner cannot do. Heuristic miner can handle big event logs.
Output: Heuristic net

When to use it: When you have real-life data with not too many different events, or when you need a Heuristic net model for further analysis in ProM
Pros: One of the advantages is that a Heuristic net can be converted to other types of process models, such as a Petri net for further analysis in ProM.
Cons:

3. Fuzzy miner : Fuzzy miner discovers process model from event data. It do not discover Petri Net with process graph. The problem with process graph is we cannot distinguish between choice and parallelism. Fuzzy miner is used to explore and get insights. It is used in commercial tools like Disco and Celonis.

Output: Fuzzy Model
When to use: When you have complex and unstructured log data, or when you want to simplify the model in an interactive manner
Pros: The Fuzzy miner uses significance/correlation metrics to interactively simplify the process model at desired level of abstraction. Compared to the Heuristic miner it can also leave out less important activities (or hide them in clusters) if you have hundreds of them.
Cons:The fuzzy model cannot be converted to other types of process modelling languages, but you can use it to animate the event log on top of the created model to get a feeling for the dynamic process behaviour.

4. Multi-phase miner: The Multi-phase miner was the first algorithm which makes use of OR split/join semantics, as found in EPCs, enabling it to express complex behaviour in relatively well-structured models. It was developed by Dr.Boudewijn van Dongen.
Output: Event-driven Process Chain (EPC)

When to use: When you have simple and structured log data and you want to export the mining result to Aris.
The Multi-phase miner folds XOR, AND, and OR connectors from so-called runs and displays the resulting model as an EPC. The EPC can then be exported to Aris

(e.g., in Aris graph format) and further processed from there.
Pros: One of the advantages of the Multi-phase miner is that it constructs a model that always fits the complete event log (more on that in a later post).
Cons: Rarely used for complex processes because the model becomes unreadable.

Types of Process mining
Three types of process mining techniques exist namely-discovery, conformance checking and enhancement. Process discovery creates process model representation using event data. Conformance checking looks the difference between real life models and discovered models. Model enhancement [6] provide improvements for existing model.
These techniques can be applied for two different types of processes-Lasanga and Spaghetti[10]. A Lasanga process is structured, regular, controllable and repetitive. A Lasanga process has a fitness value of 0.8 and hence all PM tools can be applied. It is easy and less complex in nature. Whereas Spaghetti process is unstructured, irregular, flexible, variable. Here process discovery is valuable. So we need to focus on subset of activities, subset of cases and subset of paths.
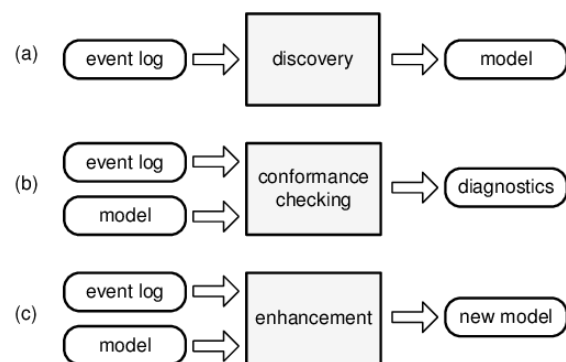

Fig.2

IV.   RESULTS AND DISCUSSION

The below given table-1 gives a comparison of various algorithms and methodologies used for process mining. The existing process mining algorithms are more suitable for mining processes without concept drift and Lasanga processes only. But due to the dynamic nature of Spaghetti process a static procedure cannot be applied to obtain process models. Concept drift in a process is identified when any two process which are sequential at a given instant of time can change from sequential to parallel and vice versa. Hence due to the existence of concept drift in a Spaghetti process we cannot achieve and analyze the required process model. Each method has both advantages and disadvantages presented in section-II of this paper.

V.   CONCLUSION AND FUTURE SCOPE

Study on various process mining approaches reveals that there is no single solution for all the problems encountered

at different stages of process discovery. Each method has pros and cons. It is the analyst based on the problem take a solution. We suggest a goal-oriented method to mine a Sphagetti process which considers entire process structure by considering domain knowledge to get an efficient goal based result before model is built. We also consider concept drift to get an efficient model for both structured and unstructured process mining to generate efficient process model which can also be used for prediction. A goal oriented mining of processes gives more insight into problem and hence the generated process model.

TABLE-1

Title	Aim	Data set used	Algorithm	Cons
Optimal process mining for large complex event logs	Complexity of event log is reduced	Heart data set	Tabu search algorithm	Factors like domain knowledge , weights on events are not considered.
A Framework for online process Concept Drift Detection from Event Streams	online concept drift detection system	Loan application data set	Online process concept drift detection algorithm	Only relationship between activities is considered other factors like time, organizational behavior are not considered**.**
Distilling Lasanga from Spaghetti Processes	To convert Spaghetti process to Lasanga process	Road accident data set	Fuzzy miner to find frequent execution paths	Analysis done on a small data set where paths and activities are not more.
A Goal-Driven Evaluation Method Based On Process Mining for Healthcare Processes	To build a Goal driven process evaluation method	Health care data set to show surgery process	Fuzzy miner and SWOT analysis	Is more suitable for health care data sets and is not efficient for other data sets because of its complexity

REFERENCES

[1] Martin Prodel, Vincent Augusto, "*Optimal process mining for large complex event logs",*2018 IEEE transactions on automation science and engineering.

[**2**] Na Liu, Jiwei Huang, LizhenCui, *"A Framework for online process Concept Drift Detection from Event Streams",* In the proceedings of 2018 IEEE International Conference on Services Computing,pp.105-113.

[3] Manoj Kumar,Likewin Thomas,Annappa, "*Distilling Lasanga from Spaghetti Processes",* In Proceedings of ACM Woodstock conference, Hong Kong, Hong Kong, March 25-27, 2017 (ISMSI '17),

[4] Wil M.P. vander Aalst ,Verlag Berlin Heidelberg, "*Process Mining Analyzing Sphagetti process"*2016 Springer pp.411-428

[5] Hind R'Bigui ,"*The state-of-the-art of business process mining challenges*" ,In proceedings of International Journal of Business Process Integration and Management 2017.

[6] Fitri Almira Yasmin et.al , *"Process Enhancement in Process Mining: A Literature Review*" , Research Gate Conference paper Dec18.

[7] Tugba Gurgen Erdogan and Ayca Tarhan A ," *Goal-Driven Evaluation Method Based On Process Mining for Healthcare Processes"* ,Appl. Sci. 2018,

[8] Josef K.J. Martens, "*Professional use of Process Mining for analyzing Business Processes",*2013-CEUR workshop proceedings .

[9] Ufuk Celik , "*Process Mining Tools Comparison",* Online Academic Journal of Information Technology 2018

[10] W.M.P. van der Aalst, "*Finding Structure in Unstructured Processes:The Case for Process Mining Lasanga"* ,In the proceedings of Application of Concurrency to System Design, ACSD 2007.

Authors Profile

Mrs.N.Swapna pursed Bachelor of Technology from Jawaharlal Nehru Tchnological University Hyderabad of Telangana,India in 2002 and Master of Technology from Jawaharlal Nehru Technological University Hyderabad in year 2011. She is currently pursuing Ph.D. and currently working as Associate Professor in Department of Computer Science, Vignana Bharathi Institute Of Technology,Hyderabad since 2013. She has published three research papers in reputed international journals and conferences which are also available online. Her main research work focuses on Process mining,Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. She had 12 years of teaching experience and 2 years of Research Experience.

Dr. L. Rama Parvathy is a Professor in the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai with 18 years of Academic Training and Teaching students including 8 years of Research. She graduated M.E. Computer Science and Engineering, from Anna University, Chennai and Ph.D. Information and Communication Engineering (I&C) from Anna University, Chennai in Computer Science and Engineering. Her research interests are Cloud Computing, Evolutionary Computing, Multi Objective Optimization and Image Processing. Her Research credential includes 12 international journal publications, two international conference publications and 10 National Conferences. She is a reviewer for reputed International Journals and Coordinator for National Conferences. She is a Subject Matter Expert (SME), Learning Assets Developer (LAD) and Trainer for Corporate companies such as HCL Technolgies, Cognizant Technology Systems.