Process Mining

NAZMUL HASAN OYON

ID: 20101528

email: nazmul.hasan.oyon@g.bracu.ac.bd

Compiled May 12, 2023

1. INTRODUCTION

Process mining is a technique for extracting useful information from data created during the execution of a business process. It enables businesses to obtain a thorough knowledge of their operations, find inefficiencies, and enhance their performance. Analyzing event logs, which collect data on the activities, resources, and interactions that occur throughout the execution of a business process, is part of the process of mining approach. Process mining may build visual representations of the process flow, highlight bottlenecks and inefficiencies, and suggest possibilities for change by analyzing these event records. Process mining has applications in a variety of industries, including banking, healthcare, logistics, and manufacturing. Process mining in finance can aid in the detection of fraudulent activities or inefficiencies in financial operations. It may be used in healthcare to improve patient care and cut wait times. It can optimize supply chain operations and discover possible process improvements in logistics and manufacturing. Typically, process mining consists of three stages: data preparation, process discovery, and process analysis. The event logs are gathered, cleaned, and translated into a suitable format for analysis during the data preparation step. Process mining techniques are used in the process discovery stage to identify the process flow, including the various activities and their interconnections. The process flow is examined at the process analysis step to find inefficiencies. To sum up, process mining can help organizations reduce costs, increase efficiency, and improve their overall performance.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

2. RELATED WORKS

Effendi and Sarno (2014) stated in the paper "Modeling Parallel Business Process Using Modified Time-Based Alpha Miner" proposes a modified version of the Alpha Miner algorithm to analyze and model parallel business processes. The modified algorithm incorporates a time-based approach that can handle events that occur concurrently in different process paths. The authors use a real-life event log from a financial institution to demonstrate the effectiveness of their modified approach in discovering parallel process models accurately and efficiently. The results show that the modified algorithm outperforms the traditional Alpha Miner algorithm and other state-of-the-art process mining techniques in terms of accuracy and efficiency. The authors conclude that the proposed modified algorithm can be a useful tool for analyzing and improving parallel business processes in various domains.

Kundra et al. (2016) propose a parallel implementation of the algorithm using both CPU and GPU architectures to improve the performance. The parallel implementation is based on the MapReduce programming model and is implemented using the Hadoop framework. And, it compares the performance of the parallel implementation on a CPU cluster and a GPU cluster. They measure the execution time, throughput, and speedup of the algorithm on both architectures. The experimental results show that the parallel implementation on the GPU architecture provides better performance than the CPU architecture. The GPU implementation achieves a speedup of up to 4.6x compared to the CPU implementation, indicating that the GPU architecture is well-suited for parallelizing the Alpha Miner algorithm. It also analyze the scalability of the parallel implementation and show that the GPU implementation can handle larger event logs than the CPU implementation, with a maximum event log size of 100,000 events for the GPU implementation compared to 50,000 events for the CPU implementation. In conclusion, the paper demonstrates the effectiveness of parallelizing the Alpha Miner algorithm using the GPU architecture, providing significant performance improvements over the CPU architecture. The results suggest that the proposed parallel implementation can scale well to larger event logs, making it suitable for real-world process mining applications.

Aalst(2012) expained an overview of the field of process mining, which involves analyzing event logs generated by information systems to discover, monitor, and improve business processes. The authors begin by highlighting the limitations of traditional methods for process modeling and analysis, which rely on subjective assessments and are often incomplete or inaccurate. They then introduce the concept of process mining, which uses data-driven techniques to automatically extract process models from event logs. There are three main types of

process mining: process discovery, conformance checking, and enhancement. Process discovery involves the automatic generation of process models from event logs, while conformance checking compares the actual process behavior with the model to identify deviations and potential problems. Enhancement aims to optimize the process performance by identifying areas for improvement based on the process model and the event log data and provide a clear and concise explanation of the key concepts and techniques used in process mining, and provide examples of real-world applications, such as fraud detection, compliance monitoring, and performance improvement.

3. METHODOLOGY

A. Data Collection

Data gathering is the initial phase of the study approach. The dataset for analysis is collected in this stage. The code imports the essential libraries for data analysis and uses the os library to import the dataset.

B. Data pre-processing

Once the data has been imported, the following step is to remove any undesirable data from the log. To filter out the process's start and end activities, the code uses the start activities filter and end activities filter methods. The variants filter function is also used to extract variations (unique sequences of activities) from the log. The case statistics function counts the number of cases in each version and sorts them descendingly. Finally, the attributes filter function is used to get the frequency of each activity in the log.

C. Data Analysis

The next step after preprocessing is data analysis. In this stage, the code analyzes the log using several techniques from the pm4py package. First, a Petri net is generated using the alpha miner technique. The Petri net depicts the loan approval process's workflow. The inductive miner technique is used to construct a process tree that also depicts the loan approval procedure's workflow.

D. Process Discovery

The following step is to extract the process model from the log. Three process discovery methods are used in the code: alpha miner, inductive miner, and heuristics miner. The alpha miner algorithm generates a Petri net, whereas the inductive miner method generates a process tree.

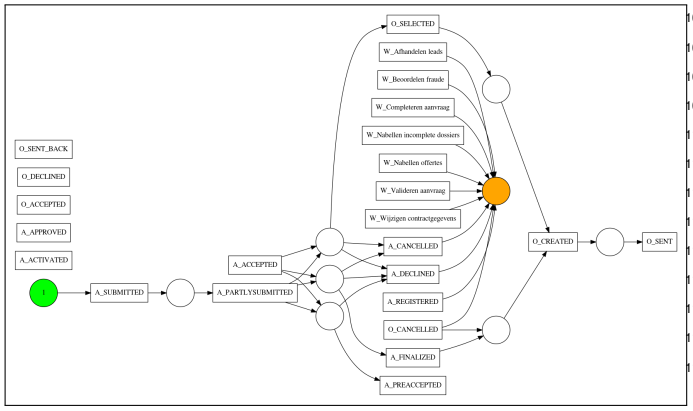

Fig. 1. Alpha Miner

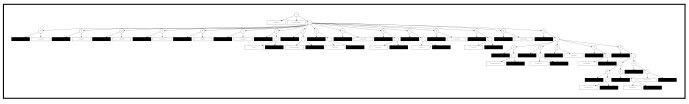

Fig. 2. Inductive Miner

A heuristic net is created using the heuristics miner algorithm. The heuristics miner method is used to create a heuristics net that depicts the most frequently utilized pathways in the loan. Finally, the directly follow graph (DFG) method is employed to build a DFG displaying the frequency of activity occurrences.

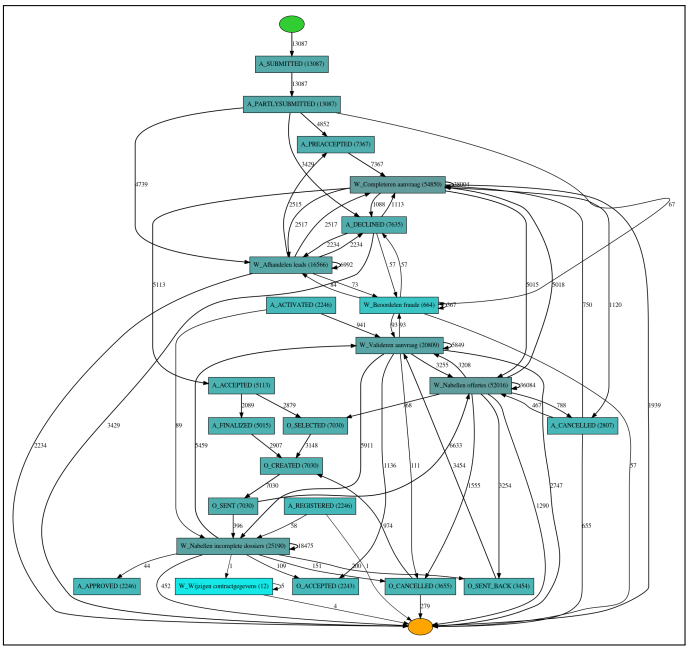

Fig. 3. Heuristic Miner

E. Conformance checking

It entails comparing the actual process to the predicted or discovered process, which is an important stage in process improvement. The goal is to discover any inconsistencies or deviations in the real process and evaluate the core reasons for these aberrations. This stage is critical for identifying bottlenecks, inefficiencies, and potential for process improvement. The process may be adjusted to become more efficient, effective, and compliant by detecting the variances. The actual process is compared to the process model, which can be a flowchart, a BPMN (Business Process Model and Notation), or any other representation of the process, to perform the conformity check. Any differences between the two processes are noted, and the underlying reasons are investigated. People, processes, technology, or external factors can all be root causes. Once the underlying reasons have been recognized, corrective actions may be implemented to address the deviations and enhance the process.

F. Process Improvement

This stage entails resolving the deviations discovered during the conformance-checking process and modifying it to make it

more efficient, effective, and compliant. The process augmentation stage may entail restructuring the process flow, automating certain processes, or enhancing communication routes between various teams or departments. The goal is to optimize the process in order to decrease cycle time, eliminate waste, and increase quality.

G. Process Monitoring

This stage entails constantly monitoring the process for new deviations and chances for improvement. Process metrics like as cycle time, throughput, failure rate, and customer satisfaction can be used to monitor the process. Organizations may guarantee that the process stays efficient and successful over time by regularly monitoring it.

H. Process Visualization

The last step is to create visual representations of the discovered process models. This is achieved using a set of visualizer functions available in the pm4py library, such as on visualizer, pt visualizer, hn visualizer, and dfg visualizer. These visualizations are helpful in gaining a better understanding of the process flow and identifying potential areas for improvement.

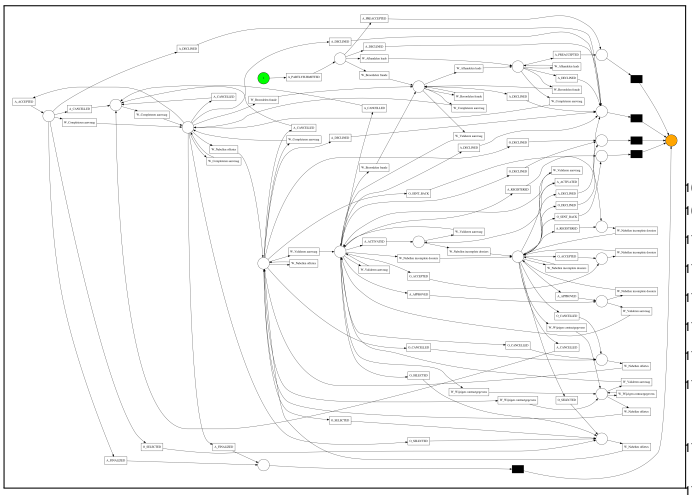

Fig. 4. Process Visualization

The Petri net visualizer (on visualizer) is used for Petri nets, while the process tree visualizer (pt visualizer) is used for process trees. The heuristic net visualizer (hn visualizer) is used for heuristic nets, and the Directly-Follows Graph visualizer (dfg visualizer) is used for visualizing the relationships between the different activities in the process. By examining these visualizations, one can identify bottlenecks and inefficiencies in the process flow and pinpoint opportunities for process improvement. The visualizations provide a clear and concise representation of the process flow, making it easier to identify areas that need attention.

4. RESULT

Four algorithms from pm4py to generate process models from the log data. These algorithms include alpha miner, inductive miner, heuristics miner, and dfg discovery. The resulting process models are visualized using the pn visualizer, pt visualizer, hn visualizer, and dfg visualization functions from pm4py. The visualizations help to identify the process flow and potential bottlenecks, inefficiencies, and opportunities for improvement.

The alpha miner algorithm generates a Petri net from the log data, while the inductive miner algorithm generates a process tree. The heuristics miner algorithm generates a heuristic net, which is another type of process model, and the dfg discovery algorithm generates a Directly-Follows Graph. The script also demonstrates how to convert the process tree generated by the inductive miner algorithm into a Petri net using the pt converter function from pm4py. The resulting visualizations provide a

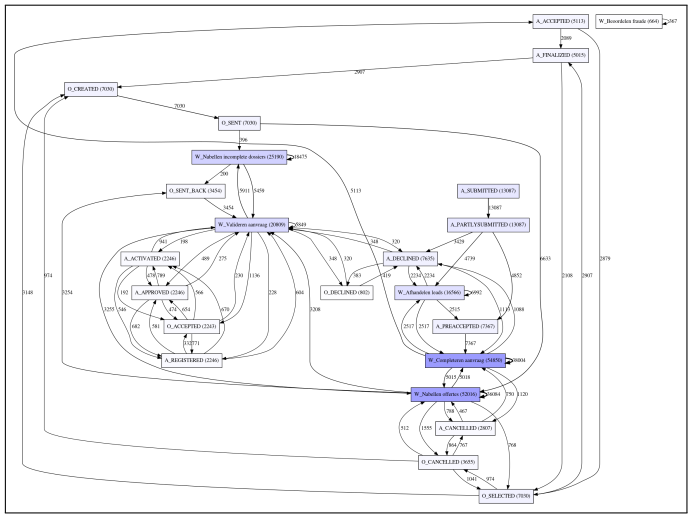

Fig. 5. Directly follow graph

clear understanding of the process flow and the frequency of each step in the process. The process models and visualizations can be used to identify inefficiencies and opportunities for improvement in the financial process being analyzed. Overall, the code provides a useful tool for conducting process mining and gaining insights into process flow and potential improvements in a financial setting.

5. DISCUSSION

Process mining is a discipline that discovers, monitors, and improves processes by using data from information systems. To extract knowledge from event logs created by information systems, it incorporates techniques from data mining, process modeling, and process analysis. One of the primary purposes of process mining is to depict real process behavior and discover inefficiencies, bottlenecks, and other issues that may be addressed. Process mining approaches have several uses, including quality assurance, compliance monitoring, risk management, and process optimization.

PM4Py is an open-source process mining platform that offers a full collection of tools for process mining activities such as process discovery, compliance verification, and process enhancement. This programme loads and processes an XES format event log containing data on personal loan applications using various process mining techniques. The log is filtered initially to extract the start and end activities, as well as the process variations. The Alpha Miner technique is then used to generate a Petri net model of the process, which is then displayed using the PM4Py package. The inductive Miner approach is also used to find a process tree, which may be transformed to a Petri net. Heuristic mining is also used to create a Heuristics net, which can aid in identifying relationships

between process activities. Finally, a Directly-Follows Graph is constructed, which can be used to identify frequently occurring sequences of activities in the process.

The fundamentals of the many process mining approaches that may be used to examine event logs. The approaches employed in this code can assist in identifying process inefficiencies and bottlenecks, which can then be further evaluated and adjusted to enhance overall process performance. To implement process mining techniques in Python, the PM4Py package offers a complete set of tools. These tools may be used to find process models, analyze process behavior, and spot process inefficiencies. This code may be modified to more complicated circumstances in order to get deeper insights into the process and acts as a starting point for additional investigation and analysis of event logs.

6. CONCLUSION AND FUTURE WORK

It appears that the study done utilizing process mining techniques has something to do with the financial sector, more especially the loan approval procedure. The example code demonstrates how process mining may be used to find bottlenecks and inefficiencies, as well as to uncover the process flow and represent it in an easily understood manner. Process mining approaches may be very helpful for process optimization, risk management, and compliance monitoring in the financial sector. Organizations may use it to better understand their processes, spot areas for development, and streamline their workflow. Process mining methods may also be used to spot possible fraud and to guarantee adherence to rules and internal standards.

As businesses continue to seek efficiency and legal compliance, process mining is expected to become even more common in the financial sector in the future. But there are other issues that must be resolved, such as poor data quality and privacy worries. Organizations must make sure they have high-quality data and adhere to privacy laws in order to deal with these issues. In order to successfully evaluate and improve their processes, businesses must also invest in process mining tools and knowledge. This has the potential to aid firms in streamlining their operations, ensuring compliance, and uncovering possible fraud. However, they must deal with issues related to data quality and privacy problems, as well as make the appropriate investments in technologies and knowledge, in order to fully enjoy the advantages.

7. REFERENCES

1. Effendi, Y. A., Sarno, R. (2018). Modeling parallel business process using modified time-based alpha miner. International Journal of Innovative Computing, Information and Control, 14(5), 1565-1582.
2. Kundra, D., Juneja, P., Sureka, A. (2016, July). Vidushi: parallel implementation of alpha miner algorithm and performance analysis on CPU and GPU architecture. In Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–September 3, 2015, Revised Papers (pp. 230-241). Cham: Springer International Publishing.
3. Van Der Aalst, W. (2012). Process mining. Communications of the ACM, 55(8), 76-83.