



PREDICTION AND RECOMMENDATION IN PROCESS MINING

B. Kamala Assistant Professor
Dept. Master of Computer Application
Sri Sairam Engineering College ,
West Tambaram, Chennai-600044.

R. Jayashree PG Scholar
Dept. Master of Computer Application
Sri Sairam Engineering College,
West Tambaram, Chennai-600044.

Abstract:

Although there has been much progress in developing process mining algorithms in recent years, no effort has been put in developing a common means of assessing the quality of the models discovered by these algorithms. the need for such an evaluation mechanism, (a) process mining researchers to compare the performance of their algorithms, and (b) end users to evaluate the validity of their process mining results. Recently, process mining emerged as a new scientific discipline on the interface between process models and event data. On the one hand, conventional Business Process Management (BPM) and Workflow Management (WfM) approaches and tools are mostly model-driven with little consideration for event data. On the other hand, Data Mining (DM), Business Intelligence (BI), and Machine Learning (ML) focus on data without considering end-to-end process models. Process mining aims to bridge the gap between BPM and WfM on the one hand and DM, BI, and ML on the other hand. Here, the challenge is to turn torrents of event data

Keywords

("Big Data") into valuable insights related to process performance and compliance. Fortunately, process mining results can be used to identify and understand bottlenecks, inefficiencies, deviations, and risks. This tutorial paper introduces basic process mining techniques that can be used for process discovery and conformance checking. Moreover, some very general decomposition results are discussed. These allow for the decomposition and distribution of process discovery and conformance checking problems, thus enabling process mining in the large. Process mining allows for the automated discovery of process models from event logs. These models provide insights and enable various types of model-based analysis. This paper demonstrates that the discovered process models can be extended with information to predict the completion time of running instances. There are many scenarios where it is useful to have reliable time predictions

Process Mining, Data Mining, Data Analysis,
Business Intelligence, Business Process Intelligence, Business Process Management, Business Process Analysis,

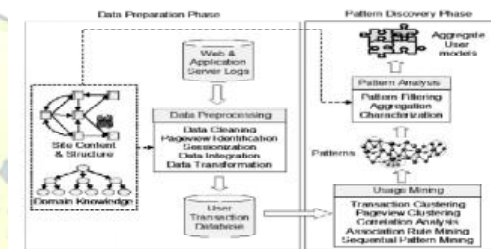
Conformance Checking,
Compliance Checking.



Introduction

Process Mining is about using smart software to continuously **analyze** and **visualize** the flow of our operations, so that we can identify specific ways to improve our business. It's easier and more cost-effective than many people realize. In **process mining** we take the data that exists in the information systems of a company and use that to visualize what is actually happening in the company's processes and how they are executed in real life. Almost all IT systems store data in **data bases** and **create logs** that can be described in process mining terms as event data. This is the basis for process mining and the **analyses conducted**. In process mining we use the event data in company IT systems to bring insight to the company's business operations. This we do by visualizing the data in process context (e.g. process flow charts) and creating analyses that give information on needed improvements and deepen the understanding of what is going on. More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts.

Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems.



Applications in process Mining

Example :

- In manufacturing industry a timely and accurate delivery to a customer is the ultimate goal.
- If a company has several factories in different regions, there usually are differences between the reliability of deliveries. It's quite easy to see they exist but harder to understand exactly where or why that is.
- Also, very often people with the loudest voice gets the most attention regardless of whether their problems actually are the



biggest. With process mining we can compare objectively the performance of different regions down to individual process steps, duration of a step, cost of a step, or the person performing the step and many more.

- Whatever event data is available in the systems it can be used in process mining.

Proposed contribution in process mining

Verification and performance analysis heavily rely on the availability of high quality models. When the **models** and **reality** have little in common, model-based analysis does not make much sense. There is often a lack of alignment between handmade models and reality. Process mining aims to address these problems by establishing a direct connection between the models and actual low-level event data about the process. Process discovery techniques allow for viewing the same reality from different angles and at different levels of abstraction. Process mining allows for the automated discovery of process models from event logs. Discovered process models can be extended with information to predict the completion time of running

instances. Annotated transition systems are used to predict the completion time. The approach has been implemented in ProM. Based on the state of the art of process mining, we can conclude that quality characteristics (failure rate metrics or loops) are poorly represented or absent in most predictive models that can be found in the literature.

Characteristics of process mining

The important characteristics of process mining:

1. **Evidence-based BPM:** Process mining is based on observed behavior recorded in event logs where intelligent techniques are used to extract knowledge.
2. **Fact-based:** Process mining is fact-based which means it is based on event data rather than opinions. The observed behaviors of the **end users** and **machines** are **recorded in logs**.
3. **Truly intelligent:** Process mining is truly intelligent meaning it learns from historic data and use the knowledge for model enhancement. It can be used to analyze the process in the system as well as can be used for conformance checking
4. **Process-centric:** Process mining is related to processes not only to



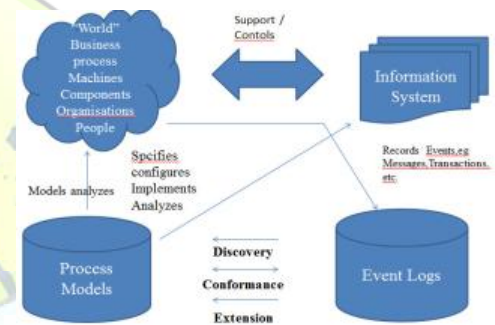
data. It is not data-centric like other mining approaches

5. **Explosion of event data:** All the activities done by people, machines and software leaves trails called event logs. More and more events are being recorded thus providing detailed information about the history processes. The high volume of data gives higher challenges of extracting useful information from it.
6. **Discovering business process:** It is always interesting to discover if the users of the system follows some processes. Process mining techniques take event logs and discover process within it if exists.
7. **Bottlenecks identifications in process:** Process mining can be used to identify bottlenecks in Information Systems by analyzing the event logs data. Using right processed data, one can find bottlenecks relating to missing steps, service interruptions or long process times. Hence, process mining can be used to identify and understand bottlenecks, inefficiencies, deviations, and risks.
8. **Conformance checking with existing model:** For quality assurance, it is often required to check if reality as recorded in log confirms to the conceptual model.

9. Recommendation of news items using process mining technique:

Process mining can be used to discover the process user reads news articles in general. These predicted process path can be recommended to new users visiting the site.

Architecture Diagram for Process Mining



- The first type of process mining is **discovery**. A discovery technique takes an event log and produces a process model without using any a-priori information. An example is the Alpha-algorithm that takes an event log and produces a process model (a Petri net) explaining the behaviour recorded in the log.
- The second type of process mining is **conformance**. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa.



- The third type of process mining is **enhancement**. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log. **Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model.** An example is the extension of a process model with performance information, e.g., showing bottlenecks.
- **Process model extension:** aims to improve a given process model based on information (e.g., time, performance, case attributes, decision rules, etc.) extracted from an event log related to the same process. There are different ways to extend a given process model with these additional perspectives, e.g., decision mining, performance analysis, and user profiling.
- **Process mining techniques:** can also be used for operational decision support activities. For instance, based on historic information, it is possible to make predictions (e.g., the remaining flow time) for running cases or to recommend suitable actions (e.g., proposing the activity that will minimize the expected costs and time). Moreover, it is possible to check, on the fly, if running cases fit with normative process models or if desired properties (defined in

Linear Temporal Logic) hold in these running cases

Process Models

Executable models may be used to force people to work in a particular manner. However, most models are not well-aligned with reality. Most hand-made models are disconnected from reality and provide only an idealized view on the processes at hand: “paper tigers”. Given (a) the interest in process models, (b) the abundance of event data, and (c) the limited quality of hand-made models, it seems worthwhile to relate event data to process models: process mining!

insight: while making a model, the modeler is triggered to view the process from various angles;

• **discussion:** the stakeholders use models to structure discussions;

• **documentation:** processes are documented for instructing people or certification purposes (cf. ISO 9000 quality management);

• **verification:** process models are analyzed to find errors in systems or procedures (e.g., potential deadlocks);

• **performance analysis:** techniques like simulation can be used to understand the factors influencing response times, service levels, etc.;

• **animation:** models enable end users to “play out” different scenarios and thus provide feedback to the designer;

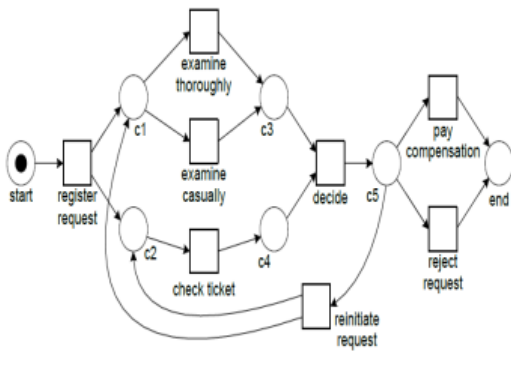
• **specification:** models can be used to describe a PAIS before it is implemented and can hence serve as a “contract”



between the developer and the end user/management; and

- configuration**: models can be used to configure a system

Example for Process Model



Process Mining by using Event logs

The first point for process mining is an event log. The event in such a log refers to an activity and is related to a particular **process instance**. Event logs can store additional data about events. Process mining techniques use supplementary information such as the resource executing or initiating the activity, **logs data**, event's time stamp, and **data attributes**. Attributes store additional information that can be used for analysis purposes. An event log contains all recorded events that relate to executed activities in a **table**. A process model is an abstraction of the real world

execution of a process. The **event logs** as a **set of events** that are mapped to the same case. The sequence of recorded events in a case is called **trace**. Process instance model describes the execution of a **single process instance**. A process model abstracts from the single behaviour of process instances and provides a model that reflects the **behaviour of all instances** that belong to the same process. Classifiers ensure the distinctness of cases and events by mapping unique names to each case and event. **Cases and events** are characterized by **classifiers** and **attributes**. First, analysts use process discovery techniques to evaluate a model from an event log. Analysts then apply conformance checking techniques to diagnose deviations between the event log and initial process model. During model creation, analysts use information from the log to repair or extend the model. They can use time stamps to add timing information such as waiting times and service times to the model. Then the resulting enhanced process model can support decision making. [6] discussed about a method, Sensor network consists of low cost battery powered nodes which is limited in power. Hence power efficient methods are needed for data gathering and aggregation

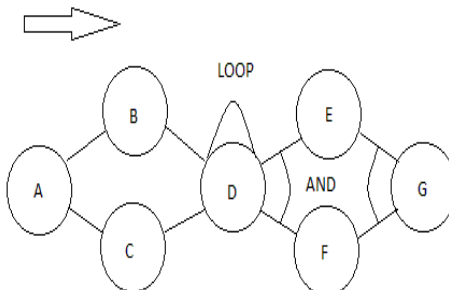


in order to achieve prolonged network life. However, there are several energy efficient routing protocols in the literature; quiet of them are centralized approaches, that is low energy conservation. This paper presents a new energy efficient routing scheme for data gathering that combine the property of minimum spanning tree and shortest path tree-based on routing schemes. The efficient routing approach used here is Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs) which is robust and localized. This is based on powerful localized structure, local minimum spanning tree (LMST). The actual routing tree is constructed over this topology. There is also a solution involved for route maintenance procedures that will be executed when a sensor node fails or a new node is added to the network.

Working of Event logs in System Methodology

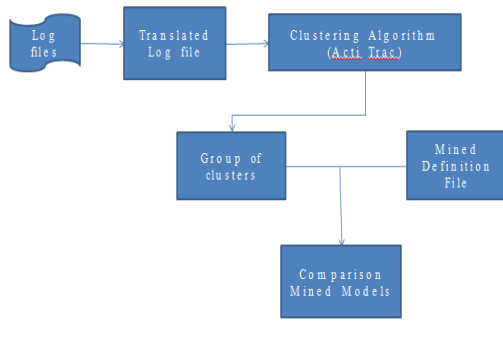
- Here take the input from user, which is a systematic event logs file. Convert this event logs file into translated tokenized logs file through the log transformation phase.
- Then the tokenized logs file is then processed through the detailed filtering process which filters the stop words, common words. On the output of filtering step we apply our rules on selection criteria's of ActiTraC algorithm on the basis of which actors, events, candidate classes, their attributes, their relationships are extracted. After that these clusters are compared to the process mining definition and generate mined process models.

Log Records:
ABDEFG, ACDDFEG, ACDFEG, ABDFEG, ACDDDEFG, ABDDEFG





Proposed Architecture of system methodology



Time prediction based on process mining:

Process mining techniques attempt to extract non-trivial and useful information from event logs. One aspect of process mining is control-flow discovery, i.e., automatically constructing a process model (e.g., a Petri net or BPMN model) describing the causal dependencies between activities. The basic idea of control-flow discovery is very simple: given an event log containing a set of traces, automatically construct a suitable process model \describing the behavior" seen in the log. Such discovered processes have proven to be very useful for the understanding, redesign, and continuous improvement of business processes. As process mining techniques are getting more mature, there is the desire to use the discovered models in an operational setting. This means that the PAIS is using results from process mining at runtime. This service, which is implemented in the process mining tool ProM, gives advice on the best possible next activity. Unlike the fixed routing in a workflow management

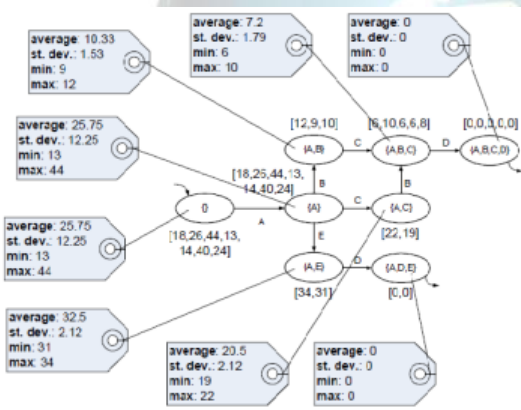
system which is strictly enforced, the recommendation service merely gives an advice what to do next. This service is using non-parametric the timestamp of the event, or data elements recorded with the event (e.g., the size of an order). Today, many information systems record such information. To explain the kind of input needed for process mining and our prediction approach has been used. Based on the event log, an annotated transition system is generated. Whenever we want to predict the completion time of some process instance, we take its partial trace (i.e., the sequence of events executed thus far) and use the state representation function l state to map the partial trace onto a state in the transition system. Using this information, a prediction is made, e.g., based on the average time to completion for earlier process instances in a similar state. As shown thus far, our approach allows for the prediction of various things including the remaining time until completion. Depending on the abstraction chosen, different predictions are possible. Moreover, under certain circumstances one prediction may be less reliable than another. When predicting the remaining time until completion. These interactions should be incorporated into the predictions.

Predictions:

Let $L _ C$ be an event log and $(S;E;T;A)$ an annotated transition system parameterized by L , l state, l event, and l measure. Moreover, let $predict \ 2 \ IB(M) \ ! \ M$ be a prediction function. For any partial trace $_N$, the predicted value is $predict(A(l \ state(_N)))$ if $l \ state(_N) \ 2 \ S$. As the above definition shows, a prediction for a running instance is made by looking up the state



corresponding to the partial trace $_N$. For this state there have been a number of measurements and using an appropriate prediction function a prediction is made. Note that it is only possible to make a meaningful prediction if the calculated state is in the annotated transition system (i.e. 1 state($_N$) 2 S) and there are a reasonable number of measurements for this state. The below diagram shows per state predictions based on predict average, predict min, predict max, predicts dev. For example, process instances that have completed activities A, B, and C (but not yet D) have an expected remaining processing time of 7:2. At the start, when no information is available yet, the predicted remaining processing is 25:75.

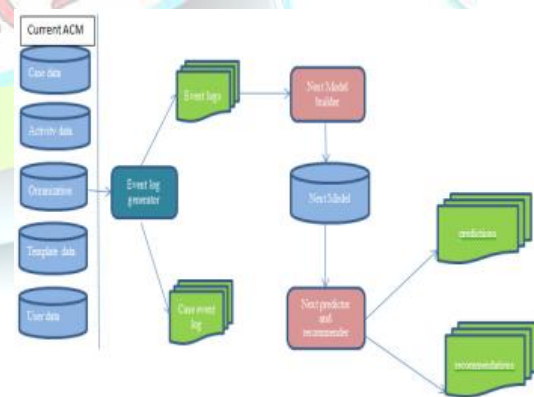


The annotated transition system showing per state: the average, standard deviation, minimum, and maximum remaining time until completion.

Next Step Recommendation and Prediction based on Process Mining in Adaptive Case Management

Process mining approaches for predictions and recommendations based on **event logs** can be integrated into existing **ACM** solutions. ACM is a new paradigm that facilitates the coordination of knowledge work through case handling. Current ACM systems **lack support** for sophisticated user guidance and for next step **recommendations** and **predictions** about the case future. However, this uses the event logs to describe how recommendation and prediction concept in **Process mining** can be applied to current ACM. The ACM uses several factors and algorithms to make prediction and recommendation unlike this thesis which considers popularity as major factor to make prediction.

Architecture Diagram for Prediction and recommendation of ACM



- The event log generator produces log entries which represent the subjects activities in a standardized XES event log format. Those logs are based on



historic events and operational data from the ACM system.

- The event logs are used by the model builder to construct underlying models based on different approaches. The model builder uses the stored activities to derive different models. Four models covering various aspects (time, deadline, decision, and goals) are created by applying different algorithms. Separate results are generated by every model, before they are combined into a composite model, called the next model.
- The next model is capable of predicting the future and recommending items based on various mining techniques. The used algorithms behind the next model are derived from data mining, statistics, and process mining.
- Generated models are used by the predictor and recommender to create predictions and recommendations for running cases. The next predictor and recommender is based on the next model and a partial trace of a currently running case. It takes all underlying models of the next models and applies the algorithms for each model.

Conclusion

Aim of this paper is to produce the appropriate information for analyzing process mining in different fields using tools and predefined software frameworks. Process Mining technique such as processmodels,event logs, petrinets, process discovery, conformance checking, bottlenecks these are used to mine the process from an information system are explained in this paper .This paper carried an extensive to design plug-ins for real-life event logs having millions of events. Given the widespread availability of process mining tools to implement new ideas and compare them with existing analysis techniques.Process discovery, event logs and Process models are evaluated using performance indicators related to speed, fitness, simplicity, generalization, and precision. In which Next Step recommendation and prediction is given based on process mining using predefined software ACM



and also prediction for business monitoring in real time environment are explained in this document.

Data Gathering “,International Journal Of Advanced Research Trends In Engineering And Technology (IJARTET), Vol. 2, Issue 10, October 2015), pp: 1-4

References:

- [1] Van der Aalst, W. M. P. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes, (1st Edition) Berlin; Heidelberg Springer.
- [2] Van der Aalst, W. M. P., Andriansyah, A., Alves de Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., Van den Brand, P., Brandtjen, R., and Buijs, J. 2016. “Process mining manifesto,” In BPM 2011 Workshops Proceedings, pp. 169–194.
- [3] Weske, M. 2012. Business process management concepts, languages, architectures, Berlin; New York: Springer
- [4] Van der Aalst, W. M. P. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes, (1st Edition) Berlin; Heidelberg:Springer.
- [5] Cook, J. E., and Wolf, A. L. 1998. “Discovering models of software processes from event-based data,” ACM Trans. Softw. Eng. Methodol. (7:3), pp. 215–249.
- [6] Christo Ananth, S.Mathu Muhila, N.Priyadharshini, G.Sudha, P.Venkateswari, H.Vishali, “A New Energy Efficient Routing Scheme for
- [7] Wil M.P. van der Aalst No Knowledge Without Processes- Process Mining as a Tool to Find Out What People and Organizations Really Do.
- [8] M. Weske. Business Process Management: Concepts, Languages, Architectures. Springer, Berlin, 2007.
- [9] Wil M.P. van der Aalst Process Mining - Discovery, Conformance and Enhancement of Business Process, 2011, SPRINGER
- [10] Wil M.P. van der Aalst, Eindhoven University of Technology, Process Mining: Overview and Opportunities