

Process Mining: Process Management with Data Mining

G.N.Singh¹ & Sandeep Aggarwal²

¹Department of Physics and Comp. Sc., Sudarshan College, Lalgaoon, INDIA

²Department of Comp. Sc. and Appl, D.A.V. College Abohar

Email: 'saggarwal_2001@yahoo.com

ABSTRACT

In this paper we propose new methods for ordering the Web pages returned from search engines. Given a few search keywords, nowadays most search engines could retrieve more than a few thousand Web pages. The problem is how to order the retrieved Web pages and then to present the most relevant Web pages first. We propose new factors to allow relevant Web pages to be ranked higher. The factors include keyword popularity, keyword to Web page popularity, and Web page popularity. The keyword to Web page popularity records which Web pages have been selected corresponding to the search keywords. The Web page popularity determines how often the Web pages have been selected and also how many popular keywords are contained in the pages. Using these popularity factors, our system is able to rank more popular pages higher, which will help most search engine users find the more popular and plausibly the more relevant pages.

Keywords: Web Page Ranking, Search Engine, Information Retrieval, Web Mining

INTRODUCTION

The Process Mining is an advanced feature of Data Mining to automatically determine and analyze actual process execution and find how the processes are performed in a completely new & process oriented way. The basic idea behind Process Mining is to extract knowledge from the event logs recorded by IT Systems. Process Mining is a process management technique that allows for the analysis of business processes based on event logs. Process mining aims at improving business processes by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs. Process Mining can be used to monitor operational processes which are like as complex workflows in a large enterprise & to illustrate complex device working. Main objective of Process Mining is to use the knowledge that is extracted from event logs to maintain business processes & to improve real business process and to redesign actual business process based on that knowledge.

- Process mining targets the automatic discovery of information from an event log. This discovered information can be used to
 - deploy new systems that support the execution of business processes;
 - or as a feedback tool that helps in auditing, analyzing and improving already enacted business processes.
- The main benefit of process mining techniques

is that information is objectively compiled. In other words, process mining techniques are helpful because they gather information about what is actually happening according to an event log of a organization, and not what people think that is happening in this organization.

- The type of data in an event log determines which perspectives of process mining can be discovered. If the log
 - (i) provides the tasks that are executed in the process and;
 - (ii) it is possible to infer their order of execution and link these tasks to individual cases (or process instances),then the control flow perspective can be mined. This is generally referred to as Process Discovery.
- If the log provides information about the persons/systems that executed the tasks, the organizational perspective can be discovered. The organizational perspective discovers information like the social network in a process, based on transfer of work, or allocation rules linked to organizational entities like roles and units.
- When the log contains more details about the tasks, like the values of data fields that the execution of a task modifies, the case perspective can be discovered.

Sources of Information for Process Mining

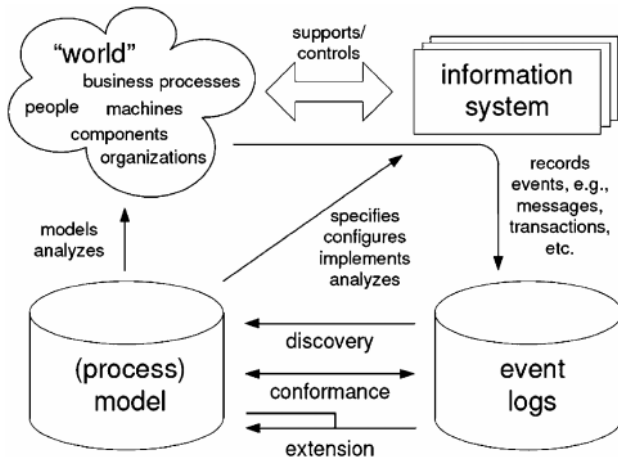


Fig. 1: Three Types of Process Mining: (1) Discovery, (2) Conformance, and (3) Extension

The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs. Clearly process mining is relevant in a setting where much flexibility is allowed or needed and therefore this is an important topic in this paper. The more ways in which people and organizations can deviate, the more variability and the more interesting it is to observe and analyze processes as they are executed. We consider three basic types of process mining (Figure 1):

1. *Discovery*: There is no a-priori model, i.e., based on an event log some model is constructed. For example, using the α -algorithm [5] a process model can be discovered based on low-level events.
2. *Conformance*: There is an a-priori model. This model is used to check if reality conforms to the model. For example, there may be a process model indicating that purchase orders of more than one million Euro require two checks. Another example is the checking of the four-eyes principle. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations.
3. *Extension*: There is an a-priori model. This model is extended with a new aspect or perspective, i.e., the goal is not to check conformance but to enrich the model. An example is the extension of a process model with performance data, i.e., some a-priori process model is used to project the bottlenecks on.

Process mining techniques are often used when no formal description of the process can be obtained by other means, or when the quality of an existing documentation is questionable. Process mining has emerged as a way to

discover or check the conformance of processes based on event logs i.e., events logged by some information system are used to extract information about activities and their causal relations. This enables organizations to learn from processes as they really take place. Also it helps to understand what is the most frequent path in a process or to what extent the cases comply with process model. Also it helps to find the throughput times of various cases, service time of a task and when the case will be completed. This information can be used to extract new information about service processes (e.g., bottlenecks, unused paths, etc.) and to check the conformance (e.g., deviations from some predefined process). Other typical problems are the presence of duplicate activities, hidden activities, non-free-choice constructs, etc. In addition, real-life logs contain noise (e.g., exceptions or incorrectly logged events) and are typically incomplete (i.e., the event logs contain only a fragment of all possible behaviors).

To understand process executed in many activities, process mining technologies are now extensively studied. However, three major problems in the current process mining techniques are identified.

1. The most process mining techniques mainly use local search strategy to generate process models.
2. Time intervals between two activities are not considered so that patterns that are different in view of time are regarded as the same behaviors.
3. No precision evaluation measure is defined to evaluate the quality of process models.

Process Mining functionality is also offered by the following commercial vendors:

- Futura Reflect, a Process Mining and Process Intelligence suite developed by Futura Technology.
- Interstage Automated Process Discovery, a Process Mining service offered by Fujitsu, Ltd. as part of the Interstage Integration Middleware Suite.
- BPM|one, offering both basic process mining functionality as well as a more comprehensive process mining module as part of the Pallas Athena BPM|one software suite.

Application Areas of Process Mining

1. Process mining can be used for:
 - Process discovery (What is the process?)
 - Delta analysis (Are we doing what was specified?)
 - Performance analysis (How can we improve?)

2. It is possible to use process mining to monitor deviations like comparing the observed events with predefined models or business rules in the context of SOX
3. With Process Mining the transaction logs of an enterprise resource planning system can be used to discover models describing processes, organizations, and products.
4. Process Mining can be used for product usage monitoring so that user actions can be observed automatically in their habitual environment. here iterative large scale user tests are feasible that helps in analyzing any deployed application and to find how does the usage behaviour relate to the opinion of the user.
5. Process mining techniques has emerged as a way to discover or check the conformance of processes based on event logs and to adapt or discover the monitor model.

Process Mining can be applied in the following fields.

- Municipalities (e.g., Harderwijk).
- Government agencies (e.g., Justice department).
- Insurance related agencies (e.g., UWV).
- Banks (e.g., ING Bank).
- Hospitals (e.g., AMC hospital).
- Multinationals (e.g., DSM, Deloitte).
- High-tech system manufacturers and their customers (e.g., Philips Healthcare, ASML).
- Media companies (e.g. Winkwaves).

Workflow Mining as Application of Process Mining

Modern business enterprises have started increasingly use the workflow paradigm to prescribe how the business process should be performed. Processes are normally represented as annotated activity graphs. The graph so produced conforms to the dependencies and past executions present in the log. This technique allow easier introduction of a workflow system and evaluation and evolution of existing process models. Recent research has focused on using a log to “reverse engineer” as workflow.

1. Workflow Mining is used to determine the structure or analyze the results of ad-hoc workflows that have no explicitly defined flow.
2. Workflow Mining help us to verify that an explicitly defined workflow is being executed as intended by the organization.
3. Workflow mining Aid in the development of explicitly defined as-is workflows by offering

suggested models that conform to the data or logs being generated.

Future Scope

Feasibility of implementation of GA's in process/workflow mining.

The use of Genetic Algorithm to overcome the problem in process/workflow mining. Genetic Algorithms have proven track record to perform efficient global search. Genetic Algorithms are robust to noise and hence the problem of event log with noise is tackled naturally. Managing the non trivial constructs can be handled by choosing gene & chromosome presentation that support them. We should experiment with real time event logs, definitions and methodology to define a suitable fitness function. We would like to propose a model that

1. That use the global search technique to generate the process models. therefore Genetic Algorithms are proposed and will be developed because they have a proven track record in global optimized search.
2. Fitness measures will be defined that are complete and precise i.e. the reproduction of all the required behavior and filtration of extra behaviour that is not required to be derived from the event log.
3. State transition with respect to the time for the active components of the pattern will be defined so that time intervals between two actives to be considered in view of the time may have the different behaviour.

REFERENCES

- [1] Process Mining: A Research Agenda by W.M.P. Vander Aalst and A.J.M.M. Weijters.
- [2] Genetic Process Mining: An Experimental Evaluation: A. K. A. de Medeiros, A. J. M. M. Weijters, W. M. P. Vander Aalst.
- [3] Application of Genetic Algorithms in Data Mining: Pankaj Nagar, Sumit Srivastava.
- [4] The Need for a Process Mining Evaluation: Framework in Research and Practice Position Paper by A. Rozinat, A.K. Alves de Medeiros, C.W. Gunther, A.J.M.M. Weijters, and W.M.P. Vander Aalst.
- [5] W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster, “Workflow Mining: Discovering Process Models from Event Logs”, *IEEE Transactions on Knowledge and Data Engineering*, **16(9)**:1128–1142, 2004.
- [6] Using Genetic Process Mining Technology to Construct a Time-Interval Process Model: Chieh-Yuan Tsai and I-Ching Chen.