

Using Process Mining to Bridge the Gap between BI and BPM

Wil van der Aalst

Eindhoven University of Technology, The Netherlands

Process mining techniques enable process-centric analytics through automated process discovery, conformance checking, and model enhancement.

The spectacular growth of digitized information makes it possible to systematically collect business-related events. The term *Business Intelligence* (BI) refers to a broad range of tools and techniques that use such event data to support decision making. Under the “BI umbrella” many three-letter acronyms have been introduced to refer to rather simple reporting and dashboard tools, e.g., Business Activity Monitoring (BAM), Corporate Performance Management (CPM), Continuous Process Improvement (CPI), and Business Process Intelligence (BPI). Only few BI tools offer mature data mining capabilities, and even the tools that do so, are not process-centric, i.e., the focus is on data and local decision making rather than end-to-end processes.

Business Process Management (BPM) techniques and tools on the other hand evolve around process models. More and more organizations are using BPM systems to support their operational processes. Process models are used to configure such systems and to analyze “as-is” and “to-be” processes. Unfortunately, these models are often completely disconnected from actual event data. Analysis results are unreliable because they are not based on observed facts, but on an idealized model of reality.

Process mining aims to bridge the gap between BI and BPM. The combination of both process models and event data allows for new forms of process-centric analytics.

Starting point for process mining is an *event log*. Each event in such a log refers to an *activity* (i.e., a well-defined step in some process) and is related to a particular *case* (i.e., a *process instance*). The events belonging to a case are *ordered* and describe one “run” of the process. Event logs may store additional information about events. In fact, whenever possible, process mining techniques use supplementary information such as the *resource* (i.e., person or device) executing or initiating the activity, the *timestamp* of the event, and other data attributes (e.g., the size of an order).

Typically, three types of process mining can be distinguished: (a) *process discovery*, (b) *conformance checking*, and (c) *model enhancement* (W. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer-Verlag, 2011). Discovery techniques learn a model from an event log without using any additional information. This results in a so-called initial process model. This model can also be made by hand. In both situations, conformance checking techniques can be used to compare the observed behavior (event logs) with the modeled behavior (initial process model). This results in diagnostics showing deviations between model and log. After conformance checking, model and log are aligned and information from the log may be used to enhance the model. The model may be repaired or extended with other perspectives such as the time or resource perspective. For example, timestamps in the event log may be used to add timing information (waiting times and service times) to the model. Such an extended model can be used for decision support.

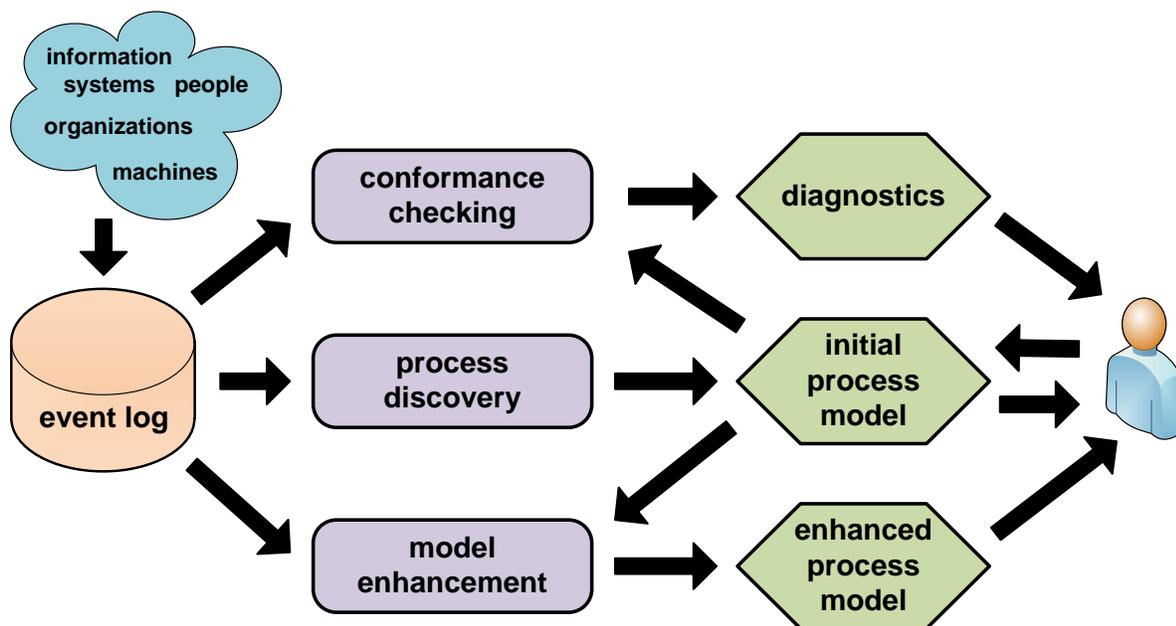


Figure 1. There are three basic types of process mining: (a) *process discovery* techniques automatically learn models from event logs, (b) *conformance checking* techniques diagnose differences between model and reality (as observed through event logs), and (c) existing models (discovered or made by hand) can also be repaired or extended using event data (*model enhancement*).

Process Discovery

Events are related to particular process instances, e.g., a blood test in a hospital is related to a patient being treated and a payment transaction in a sales process is related to a particular customer order. All events related to a *particular* process instance (i.e., case) can be ordered. Since these events also refer to activities, we can describe each case as a trace of activity names.

For a process with activities *A*, *B*, *C*, *D*, and *E*, we may find the following traces: *ABCE*, *ACBE*, and *ADE*. For example, an event log may contain information about 238 cases following trace *ABCE*, 56 cases following trace *ACBE*, and 88 cases following trace *ADE*. Process discovery algorithms can transform such an event log into a process model that adequately describes the observed behavior. For this simple example it is easy to construct a process that always starts with activity *A*, ends with activity *E*, and in-between either *B* and *C* occur (in any order) or just *D* occurs. For processes consisting of dozens or even hundreds of activities, this is much more challenging.

Figure 2 shows a fragment of a larger event log from a claim handling process. Based on such an event log we can learn the control-flow model showing the ordering of activities. The discovered model depicted in Figure 2 is expressed in the so-called BPMN notation and nicely illustrates that concurrency, choices, loops, and other control-flow constructs can be learned from example traces. As shown, the process always starts with activity *register request* and ends with *pay compensation* or *reject request*. The examination and the checking of the ticket can be done concurrently. There are two types of examinations. The decision to pay or to reject is based on the examination and check activities. It is also possible that no decision can be made yet. Activity *reinitiate request* restarts the examination and check activities thus modeling a loop construct.

In recent years, dozens of process discovery algorithms have been proposed. They are able to extract models from a wide variety of events logs in different domains (banking, insurance, manufacturing, high-tech systems, e-government, e-learning, logistics, and healthcare).

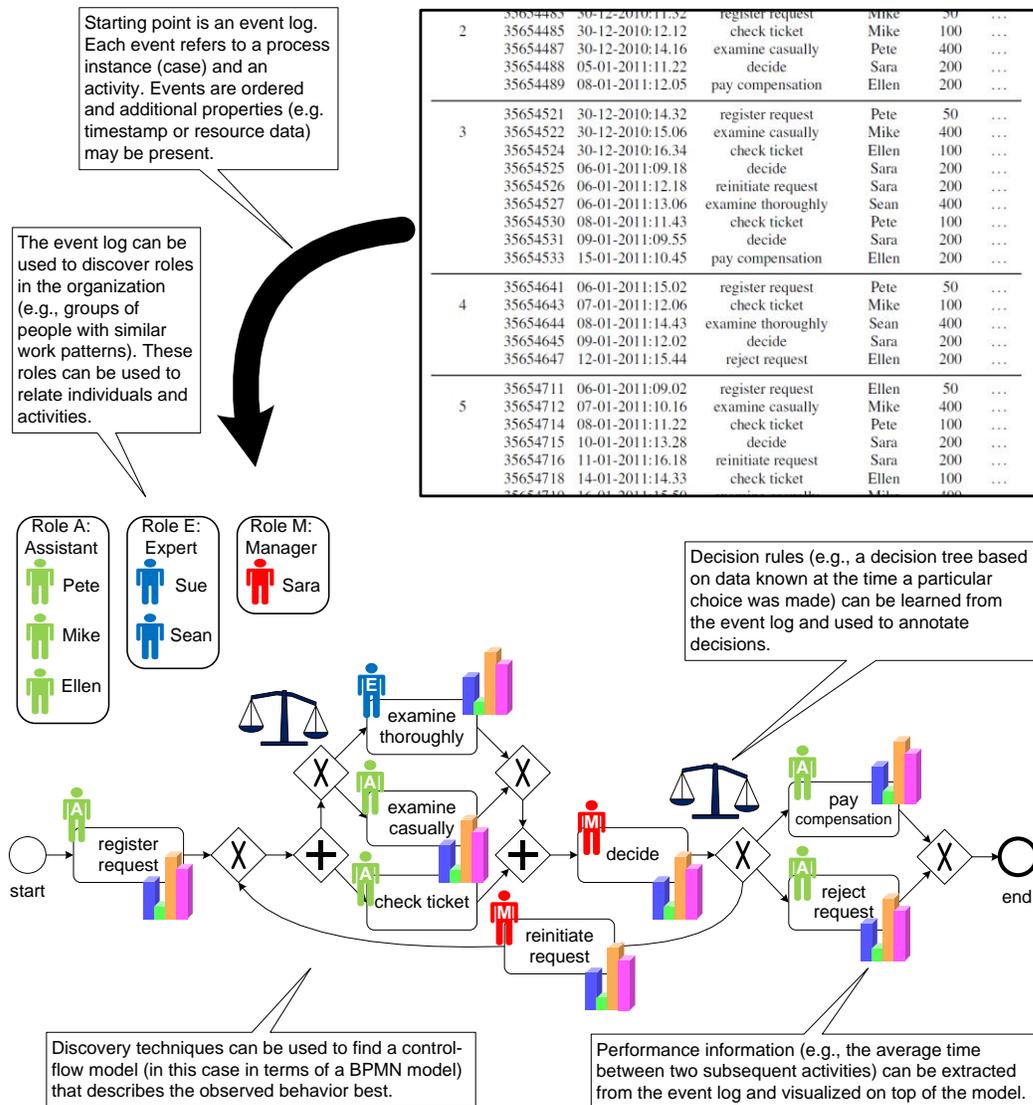


Figure 2. Overview of process mining showing the different process mining perspectives (control-flow, time, resource, and data) extracted from the event log.

Conformance Checking

Process models may be *descriptive* (showing what really happens) or *normative* (defining what should happen) and can be made by hand or discovered through process mining. In all cases it is interesting to compare model and reality (as recorded in the event log). Conformance checking techniques can be used to discover discrepancies between the modeled behavior and the observed behavior. These techniques provide metrics for the degree of conformance and diagnostic information explaining the observed differences. Moreover, it is possible to drill down and apply process discovery techniques to the non-conforming cases.

Conformance checking can be used to judge the quality of discovered process models. However, more important, it can also be employed as an enabling technology for auditing, six sigma, and compliance checking.

Model Enhancement

The third type of process mining also uses a model and an event log as input. However, now the model is improved or extended. For example, Figure 2 illustrates how a process model can be extended using timestamp information in the event log. Timestamps of causally related events can be used to measure durations between two subsequent activities. For example, analysis may show that it takes on average 21 days to make a decision after checking the ticket. This information can be used to show bottlenecks and predict remaining flow times for running cases.

If the event log contains information about resources, it is also possible to discover roles, work distribution mechanisms, and resource characteristics. Additional event and case attributes can also be used to learn decision rules explaining the choices made in the process. For example, one may learn that cases that are thoroughly checked by Sue tend to be rejected.

More information about process mining (slides, articles, software, example logs, etc.) can be obtained from www.processmining.org.

Process Mining Manifesto <<start text box>>

The growing interest in log-based process analysis motivated the establishment of the *IEEE Task Force on Process Mining* in 1999 (<http://www.win.tue.nl/ieeetfpm/>). The goal of this task force is to promote the research, development, education, and understanding of process mining. Members of the task force include representatives of more than a dozen commercial software vendors (e.g., Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Fujitsu, Infosys, and Fluxicon), ten consultancy firms (e.g., Gartner and Deloitte), and over twenty universities.

Recently, the task force released a *Process Mining Manifesto* describing six *guiding principles* and *eleven challenges*. The manifesto is supported by 53 organizations and 77 process mining experts contributed to it. The active contributions from end-users, tool vendors, consultants, analysts, and researchers illustrate the growing significance of process mining as a bridge between BI and BPM.

The guiding principles in the manifesto describe best practices in process mining. For example, the fourth guiding principle states that “Events Should Be Related to Model Elements”. This principle emphasizes the importance of relating the event log to the model. As illustrated by Figure 2, process discovery is just the starting point for process

analysis and improvement. After relating events to model elements, it is possible to *replay* the event log on the model. Replay may be used to reveal discrepancies between an event log and a model, e.g., some events in the log may be impossible according to the model. Timestamps in the event log can be used to analyze the temporal behavior. For example, bottlenecks identified during relay may be used for reengineering purposes or for making predictions about currently running cases.

Despite the applicability of existing process mining techniques and tools, there are still many challenging open problems. The manifesto lists eleven challenges. One example is “Dealing with Concept Drift in Process Mining”. This challenge refers to the problem that processes may change over time due to periodic/seasonal changes (“in December there is more demand”), changing economic conditions, or new laws and regulations. Such changes may significantly impact the performance of a process. Therefore, it is vital to detect and analyze concept drift.

The manifesto describes the guiding principles and challenges in detail and can be obtained from <http://www.win.tue.nl/ieeetfpm/>.

Google Maps for Business Processes

Process mining is an important tool for modern organizations that need to manage non-trivial operational processes. On the one hand, there is an incredible growth of event data. On the other hand, processes and information need to be aligned perfectly in order to meet requirements related to compliance, efficiency, and customer service. Process mining techniques can help to achieve such goals.

The principles and challenges described in the manifesto illustrate that process mining is a new and exciting technology. The ultimate goal is to provide organizations with “Google Maps functionality” for their operational business processes. At any point in time, there should be an up-to-date map for each process. Such process maps must be tailored towards the intended use. It should be possible to seamlessly zoom in and out. When zooming out less important activities and paths should disappear or amalgamate into aggregate nodes like in Google Maps. It should also be possible to project real-time information on such process maps. This way, information systems can visualize “traffic jams” in processes and suggest alternative routes for delayed cases.

Unlike traditional approaches the goal is not to construct a single static model. Process mining techniques can be used to dynamically generate process maps based on the most recent data and tailored towards the questions that need to be answered.

Wil van der Aalst is a professor of information systems at Eindhoven University of Technology (The Netherlands) and Queensland University of Technology (Australia). Contact him at w.m.p.v.d.aalst@tue.nl