Organizations often construct business process models to document how they think processes are executed (as-is) or should be executed (to-be). But are these models accurate? Do they represent real process execution or do they only reflect people’s perceptions about processes? Typically, an auditor has to pick samples to verify certain regulations. These sample cases are mostly selected rather randomly. But isn’t it possible to use a technique that directs you quickly to suspicious cases? Wouldn’t it be interesting to be able to focus on exceptional cases right from the start? Process mining helps you with that. It is an umbrella term for dozens of techniques that derive interesting information from historical process data that can be found in so-called log files of your software systems. This will empower an auditor with the tools to focus on real process execution and to quickly jump to exceptional behavior.

**INTRODUCTION**

In a plane, every event is recorded in a so-called black box. On a ship, the captain notes every event in a log book. In software, something similar occurs: debug or audit information is stored in files or databases. This way, enormous amounts of data about process executions are available, but it is mostly left untouched. Process mining attempts to use this particular data effectively to address three practical questions.

1. How is a business process executed in reality (process discovery)?
2. Which deviations from the normal or desired behavior (conformance checking) exist?
3. How to enrich a process model, for example with performance data or decision determinants (process enhancement)?

**COLLECTING DATA**

The recorded process data from various IT systems (e.g., ERP software, CRM software, accounting software) first needs to be hierarchically structured to assemble an event log. This event log forms the base for all process mining techniques. Figure 1 shows an example of an event log of a customer support process. Another typical example of data used to construct the log are status updates on certain fields.

Truly, the search for and preparation of data is the most challenging part of a process mining effort, because it depends on your particular information system setup. Once the data is covered in an event log, several process mining techniques can be applied.

An event log is a file that captures the history of executed processes. This history is constructed in traces and events. A trace is a group of activities (events) that belong to the same process execution (e.g., all events for the same customer, for the same sales order, for the
same ordered product, ...]. In a first phase you decide which view you wish to take on the process [e.g., do you wish to follow customers throughout the process, or sales orders, or ordered products, ... ?]. Having selected the case you wish to follow throughout a process, the traces of activities can be created. A trace contains all the activities that are executed on the case you follow, like ‘Creating a purchase order’, ‘Releasing a purchase order’ etc in a procurement process. Aside from the activity traces, the log can contain other interesting attributes (e.g., the customer number, ...). To create an event log the minimal information you need for events is:

- a case id
- an event name
- a time stamp

Making use of this information, traces can be defined (through the case id) and the event log can be constructed.

**PROCESS MINING TECHNIQUES**

After constructing the event log, various process mining techniques are accessible. They can be seen as mathematical pattern seeking algorithms. Based on reocurring patterns, the main and exceptional ways of executing a process are discovered and represented in a process model.

Process discovery [see Figure 2] is only concerned about detecting how processes were executed in reality. For an event log, you basically select a specific discovery technique to form the generated process model based on actual process data. The model can be examined to explore exceptional paths. Diverse regulations such as Sarbanes-Oxley Act demand proper and accurate business process documentation. With process mining you can be confident the process models are up-to-date.

Conformance checking can be used to check if certain rules are violated [e.g., segregation of duties]. It instantly gives an overview of all violations. The input for such an analysis is an event log and a set of business rules to check compliance for.
**Decision mining** is a specific process enhancement technique that searches for determinants that drive decisions in the process. It can be used to **search for causes** to select inappropriate paths in the process. In which period do these exceptions occur? Or is there a certain group of users that need more extensive training? Maybe the inadequate steps are only performed for larger orders? The more attributes you add to the event log, the more information can be considered to find the determinants.

Another typical example of process enhancement is **performance analysis**. For this technique, each task needs to be represented by two moments in the log: a starting moment and an ending moment. This way a difference is made between effective process time and waiting time.

**Bottlenecks** can be easily exposed by excessive waiting times, but process mining also offers techniques for **social network mining**. These techniques explore how people (or devices) work together. Bottlenecks are typically involved in (too) many cases or tasks of a process. Social network analysis allows to look at the impact of bottlenecks: Which other people/devices/tasks are involved?

### Analysis of Results

Process mining comes with some explicit limitations. First, the only goal of process mining techniques is to present useful information based on actual process data. It is still up to analysts to decide what the discovered information means, what needs to be changed to solve some issues (to-be), etc. It must be seen as an additional instrument in the toolset of managers and (process) analysts to determine the current situation of processes (as-is).

Next, as with all mining techniques the quality of results heavily depends on the input. While much software packages store lots of data, it might be difficult to filter the right information, at the right granularity level. It is also not certain that every single step in a process is reflected in the data. This is not a big issue, but it deserves explicit attention to mention that one has to keep this fact in mind when interpreting process mining outcomes.

### Steps in a Typical Project

A typical process mining project for audit consists of seven steps:

- interview
- data dump
- event log creation
- process mining
- audit discussion
- process mining 2
- reporting

As most data mining techniques, process mining techniques search for answers to some typical questions. The first step in a project is to formulate those questions for your organization. Are you concerned about optimizing performance (e.g., in a lean/six sigma exercise)? Are you concerned about finding violations against predefined policies? Are you concerned about comparing process executions of different business units?

Third step is the creation of the event log from the collected data. For this step, the extracted data of the information system must be reorganized in the desired format. All information in the related tables must be collected in one single table in the right format. Once this table is created, this should be converted into a minable file. Tools exist to convert an .xls or .csv file to the appropriate format.

As a fourth step different explorative process mining techniques will be applied to the event log. Here are some examples:

- basic log information (number of cases, of events, events per case, etc.) See Figure 3.
- top ten most frequent process variations
- cases with the longest lead time
- general process model discovery
- simple segregation of duties check
- testing other internal controls
- social network analysis

In an audit discussion first results are discussed and new, more specific question are raised. A second or perhaps third round of process mining should provide the missing answers and finally all conclusions are bundled in an audit report.

Read more about process mining in the Process Mining Manifesto that can be downloaded from http://www.win.tue.nl/ieeetfpm. Please don’t hesitate to contact the authors if you would like to receive more information.

Jan Claes  
Teaching assistant, Ph. D. Student  
Ghent University  
jan.claes@ugent.be

Mieke Jans  
Deloitte Bedrijfsrevisoren  
Berkenlaan 8B  
1831 Diegem  
mjans@deloitte.com

Screenshots from Fluxicon Disco  
http://fluxicon.com/disco/