Path Mining and Process Mining for Workflow Management Systems

Jorge Cardoso
SAP AG, Germany

W.M.P. van der Aalst
Eindhoven University of Technology, The Netherlands

INTRODUCTION

Business process management systems (Smith and Fingar 2003) provide a fundamental infrastructure to define and manage business processes and workflows. These systems are often called process aware information systems (Dumas, Aalst et al. 2005) since they coordinate the automation of interconnected tasks. Well-known systems include Tibco, WebSphere MQ Workflow, FileNet, COSA, etc. Other types of systems, such as ERP, CRM, SCM, and B2B, are also driven by explicit process models and are configured on the basis of a workflow model specifying the order in which tasks need to be executed.

When process models or workflows are executed, the underlying management system generates data describing the activities being carried out which is stored in a log file. This log of data can be used to discover and extract knowledge about the execution and structure of processes. The goal of process mining is to extract information about processes from logs.

When observing recent developments with respect to *process aware information systems* (Dumas, Aalst et al. 2005) three trends can be identified. First of all, workflow technology is being embedded in service oriented architectures. Second, there is a trend towards providing more flexibility. It is obvious that in the end business processes interface with people. Traditional workflow solutions expect the people to adapt to the system. However, it is clear that in many situations this is not acceptable. Therefore, systems are becoming more flexible and adaptable. The third trend is the omnipresence of event logs in today's systems. Current systems ranging from cross-organizational systems to embedded systems provide detailed event logs. In a service oriented architecture events can be monitored in various ways. Moreover, physical devices start to record events. Already today many professional systems (X-ray machines, wafer stepper, high-end copiers, etc.) are connected to the internet. For example, Philips Medical Systems is able to monitor all events taking place in their X-ray machines.

The three trends mentioned above are important enablers for path mining and process mining. The abundance of recorded events in structured format is an important enabler for the analysis of run-time behavior. Moreover, the desire to be flexible and adaptable also triggers the need for monitoring. If processes are not enforced by some system, it is relevant to find out what is actually happening, e.g., how frequently do people deviate from the default procedure.

BACKGROUND

Path mining can be seen as a tool in the context of Business Process Intelligence (BPI). This approach to path mining uses generic mining tools to extract implicit rules that govern the path of tasks followed during the execution of a process. Generally, the realization of a process can be carried out by executing a subset of tasks. Path mining is fundamentally about identifying the subset of tasks that will be potentially be triggered during the realization of a process. Path mining is important to process Quality of Service (QoS) prediction algorithms (Cardoso, Miller et al. 2004). In processes for e-commerce, suppliers and customers define a contract between the two parties, specifying QoS items such as products or services to be delivered, deadlines, quality of products, and cost of services. A process, which typically has a graph-like representation, includes a number of linearly independent control paths (i.e. paths that are

executed in parallel). Depending on the path followed during the execution of a process, the QoS may substantially be different. If we can predict with a certain degree of confidence the path that will be followed at runtime, we can significantly increase the precision of QoS estimation algorithms for processes.

Process mining has emerged as a way to analyze systems and their actual use based on the event logs they produce (Aalst, Dongen, et al. 2003; Aalst, Weijters, Maruster, 2004; Aalst, Reijers, et al. 2007). Note that, unlike classical data mining, the focus of process mining is on concurrent processes and not on static or mainly sequential structures. Process mining techniques attempt to extract non-trivial and useful information from event logs. One element of process mining is control-flow discovery, i.e., automatically constructing a process model (e.g., a Petri net) describing the causal dependencies between activities. In many domains, processes are evolving and people, typically, have an oversimplified and incorrect view on the actual business processes. Therefore, it is interesting to compare reality (as recorded in the log) with models. Since process mining is so important for organizations, there was the need to develop a system which implements the most significant algorithms developed up to date. Therefore, the ProM framework has been developed as a completely plug-able environment. Different research groups spread out over the world have contributed to ProM. Currently there are more than 150 plug-ins available, thus supporting all aspects of process mining (Aalst, Reijers, et al. 2007).

SETTINGS

This section presents a typical business process model and illustrates also a typical process log. These two elements will be use to explain the concepts of path mining and process mining in the next section.

Business Process Scenario

A major bank has realized that to be competitive and efficient it must adopt a new and modern information system infrastructure. Therefore, a first step was taken in that direction with the adoption of a workflow management system to support its business processes. All the services available to customers are stored and executed under the supervision of the workflow system. One of the services supplied by the bank is the loan process depicted in Figure 1.

The process of the scenario is composed of fourteen tasks. For example, The Fill Loan Request task allows clients to request a loan from the bank. In this step, the client is asked to fill in an electronic form with personal information and data describing the condition of the loan being requested. The second task, Check Loan Type, determines the type of loan a client has requested

Figure 1. The loan process