Decision Support Systems 46 (2008) 300–317

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Towards comprehensive support for organizational mining

Minseok Song *, Wil M.P. van der Aalst [1]

Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 11 May 2007
Received in revised form 10 June 2008
Accepted 3 July 2008
Available online 11 July 2008

Keywords:
Process mining
Social network analysis
Business process management
Workflow management
Data mining
Petri nets

ABSTRACT

Process mining has emerged as a way to analyze processes based on the event logs of the systems that support them. Today's information systems (e.g., ERP systems) log all kinds of events. Moreover, also embedded systems (e.g., medical equipment, copiers, and other high-tech systems) start producing detailed event logs. The omnipresence of event logs is an important enabler for process mining. The primary goal of process mining is to extract knowledge from these logs and use it for a detailed analysis of reality. Lion's share of the efforts in this domain has been devoted to *control-flow discovery*. Many algorithms have been proposed to construct a process model based on an analysis of the event sequences observed in the log. As a result, other aspects have been neglected, e.g., the organizational setting and interactions among coworkers. Therefore, we focus on *organizational mining*. We will present techniques to discover organizational models and social networks and show how these models can assist in improving the underlying processes. To do this, we present new process mining techniques but also use existing techniques in an innovative manner. The approach has been implemented in the context of the ProM framework and has been applied in various case studies. In this paper, we demonstrate the applicability of our techniques by analyzing the logs of a municipality in the Netherlands.

1. Introduction

Business Process Management (BPM) systems provide a broad range of facilities to enact and manage operational business processes. Ideally, these systems should provide support for the complete BPM life-cycle: (re)design, configuration, execution, control, and diagnosis of processes. However, existing BPM tools are unable to support the full life-cycle [17]. There are clearly gaps between the various phases (i.e., users need to transfer or interpret information without any support) and some of the phases (e.g., the redesign and diagnosis phases) are not supported satisfactorily.

Process mining techniques can be used to support the redesign and diagnosis phases by analyzing the processes as they are being executed. Process mining can be seen in the broader context of Business (Process) Intelligence (BI) and Business Activity Monitoring (BAM). Commercial BI and BAM tools are not doing any process mining. They typically look at aggregated data seen from an external perspective (frequencies, averages, utilization, service levels, etc.). Unlike BI and BAM tools, process mining looks "inside the process" (What are the causal dependencies?, Where is the bottleneck?, etc.) and at a very refined level. In the context of a hospital, BI tools focus on performance indicators such as the number of knee operations, the length of waiting lists, and the success rate of surgery. Process mining is more concerned with the paths followed by individual patients and whether certain procedures are followed or not.

Process mining requires the availability of an event log. Luckily, event logs are widely available today and the total volume of events being recorded is still growing at a spectacular rate. Events logs may originate from all kinds of systems ranging from enterprise information systems to embedded systems. Process mining is a very broad area both in terms of applications (from hospitals and banks to embedded systems in cars, copiers, and sensor networks). Most of the process mining research has been focusing on *control-flow*

* Corresponding author. Tel.: +31 40 247 3205; fax: +31 40 243 2612.
E-mail addresses: m.s.song@tue.nl (M. Song), w.m.p.v.d.aalst@tue.nl (W.M.P. van der Aalst).
[1] Tel.: +31 40 247 4295; fax: +31 40 246 3992.

0167-9236/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2008.07.002

discovery, i.e., constructing a process model based on an event log while other aspects have been neglected, e.g., the organizational setting and interactions among coworkers.

The focus of this paper is on *organizational mining*. The observation that human behavior is highly relevant for the performance of processes, suggests that comprehensive support for this is needed. Process mining is most interesting in situations where processes are not completely controlled by systems. This is of course the case in any environment where humans play a dominant role. For example, in a hospital and many other professional organizations, processes "emerge" because of human decision making. The discovery of organizational knowledge, such as organizational structures and social networks, enables managers to understand organizational structures and improve business processes. Therefore, organizational mining assists in understanding and improving organizational and social structures. For example, social networks show the communication structures in enterprises. This can be used to design communication infrastructures or office layouts.

In this paper, we describe the challenges related to organizational mining and try to address them in a comprehensive manner. We elaborate issues in organizational mining and distinguish three types of organizational mining (1) *Organizational model mining*, (2) *Social network analysis*, and (3) *Information flows between organizational entities*. For organizational model mining, we explain four kinds of methods and their characteristics. For the social network analysis, we summarize our previous approach [28] and explain the applicability of the social network analysis. A method to derive organizational entities from social networks is also proposed. Our process mining tool (ProM) supports the methods proposed in this paper.

The remainder of this paper is organized as follows. We provide an overview of process mining and organizational mining in Section 2. Section 3 presents a simple example process that is used throughout this paper. Then, Section 4 introduces important notions such as process log and organizational model in much more detail. Section 5 explains the organizational mining methods along with an example. Section 6 describes the case study which demonstrates the applicability of our approach. Section 7 reviews related work. Finally, Section 8 concludes the paper.

2. Process mining

Process mining is applicable to a wide range of systems. These systems may be pure information systems (e.g., ERP systems) or systems where the hardware plays a more prominent role (e.g., embedded systems). The only requirement is that the system produces *event logs* thus recording (parts of) the actual behavior.

An interesting class of information systems that produce event logs are the so-called *Process-Aware Information Systems* (PAISs) [9]. Examples are classical workflow management systems (e.g. Staffware), ERP systems (e.g. SAP), case handling systems (e.g. FLOWer), PDM systems (e.g. Windchill), CRM systems (e.g. Microsoft Dynamics CRM), middleware (e.g., IBM's WebSphere), hospital information systems (e.g., Chipsoft), etc. These systems provide very detailed information about the activities that have been executed.

This section first provides an overview of process mining and the focuses on organizational mining.

2.1. Overview of process mining

The goal of process mining is to extract information (e.g., process or organizational models) from these logs, i.e., process mining describes a family of a-posteriori analysis techniques exploiting the information recorded in the event logs. Typically, these approaches assume that it is possible to sequentially record events such that each event refers to an activity (i.e., a well-defined step in the process) and is related to a particular case (i.e., a process instance). Furthermore, some mining techniques use additional information such as the performer or originator of the event (i.e., the person/resource executing or initiating the activity), the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Process mining addresses the problem that most "process/system owners" have limited information about what is actually happening. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what *actually* happens. Only a concise assessment of reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in system or process redesign efforts.

The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs. As shown in Fig. 1, we consider three basic types of process mining: (1) *discovery*, (2) *conformance*, and (3) *extension*.

Traditionally, process mining has been focusing on *discovery*, i.e., deriving information about the original process model, the organizational context, and execution properties from enactment logs. There is no a-priori model, i.e., based on an event log some model is constructed. An example of a technique addressing the control flow perspective is the α-algorithm, which constructs a Petri net model [18] describing the behavior observed in the event log. However, process mining is not limited to process models (i.e., control flow) and recent process mining techniques are more and more focusing on other perspectives, e.g., the organizational perspective or the case perspective. For example, there are approaches to extract social networks from event logs and

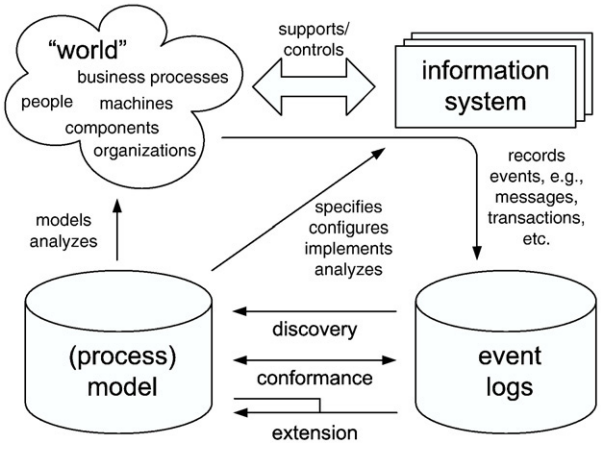

Fig. 1. The process mining overview.