How People Really (Like To) Work

Wil Aalst

How People Really (Like To) Work
Comparative Process Mining To Unravel Human Behavior

Wil M.P. van der Aalst

Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands. w.m.p.v.d.aalst@tue.nl

Abstract. Software forms an integral part of the most complex artifacts built by humans. Communication, production, distribution, healthcare, transportation, banking, education, entertainment, government, and trade all increasingly rely on systems driven by software. Such systems may be used in ways not anticipated at design time as the context in which they operate is constantly changing and humans may interact with them an unpredictable manner. However, at the same time, we are able to collect unprecedented collections of event data describing what people and organizations are *actually* doing. Recent developments in process mining make it possible to analyze such event data, thereby focusing on *behavior* rather than correlations and simplistic performance indicators. For example, event logs can be used to automatically learn end-to-end process models. Next to the automated discovery of the real underlying process, there are process mining techniques to analyze bottlenecks, to uncover hidden inefficiencies, to check compliance, to explain deviations, to predict performance, and to guide users towards "better" processes. Process mining reveals how people really work and often reveals what they would really like to do. Event-based analysis may reveal workarounds and remarkable differences between people and organizations. This keynote paper highlights current research on *comparative process mining*. One can compare event data with normative process models and see where people deviate. Some of these deviations may be positive and one can learn from them. Other deviations may reveal inefficiencies, design flaws, or even fraudulent behavior. One can also use *process cubes* to compare different systems or groups of people. Through slicing, dicing, rolling-up, and drilling-down we can view event data from different angles and produce process mining results that can be compared.

1 Events are Everywhere!

The term "Big Data" is often used to refer to the incredible growth of data in recent years. However, the ultimate goal is not to collect more data, but to turn data into real value. This means that data should be used to improve existing products, processes and services, or enable new ones. This explains the need for more *data scientists* [3]. A data scientist should be able to answer questions of the kind: *What happened?*, *Why did it happen?*, *What will happen?*, and *What is the best that can happen?* [3]. These questions all refer to the *behavior* of people, organizations, and systems. Hence, we consider *event data* to be most important source of information.

1.1 Internet of Events

Events may take place inside a machine (e.g., an X-ray machine or baggage handling system), inside an enterprise information system (e.g., a order placed by a customer), inside a hospital (e.g., the analysis of a blood sample), inside a social network (e.g., exchanging e-mails or twitter messages), inside a transportation system (e.g., checking in, buying a ticket, or passing through a toll booth), etc. In all of the above examples, software is instrumented to record events. These events tell us how people and organizations behave and use the systems at their disposal.

In [3], we coined the term the *Internet of Events* (IoE) to refer to all event data available. The IoE is composed of:

- The *Internet of Content* (IoC): all information created by humans to increase knowledge on particular subjects. The IoC includes traditional web pages, articles, encyclopedia like Wikipedia, YouTube, e-books, newsfeeds, etc.
- The *Internet of People* (IoP): all data related to social interaction. The IoP includes e-mail, facebook, twitter, forums, LinkedIn, etc.
- The *Internet of Things* (IoT): all physical objects connected to the network. The IoT includes all things that have a unique id and a presence in an internet-like structure. Things may have an internet connection or tagged using Radio-Frequency Identification (RFID), Near Field Communication (NFC), etc.
- The *Internet of Locations* (IoL): refers to all data that have a spatial dimension. With the uptake of mobile devices (e.g., smartphones) more and more events have geospatial attributes.

The above sources of event data not only reflect the abundance of event data, they also illustrate our reliance on complex software artifacts. Software forms an integral part of the most complex artifacts built by humans. Software systems may comprise hundreds of millions of program statements, written by thousands of different programmers, spanning several decades. Their complexity surpasses the comprehensive abilities of any single, individual human being. Moreover, *software must run in an ever changing context composed of different software components, different hardware configurations, may be applied in ways not anticipated at design time.* Classical modeling approaches have failed to cope with this complexity. This makes it essential to learn from systems "in vivo". We can only learn how people use systems by observing them both in their natural habitat. The event data that are omnipresent make this possible.

1.2 Event Logs

Process mining provides a powerful way to analyze operational processes based on event data. Unlike classical purely model-based approaches, process mining

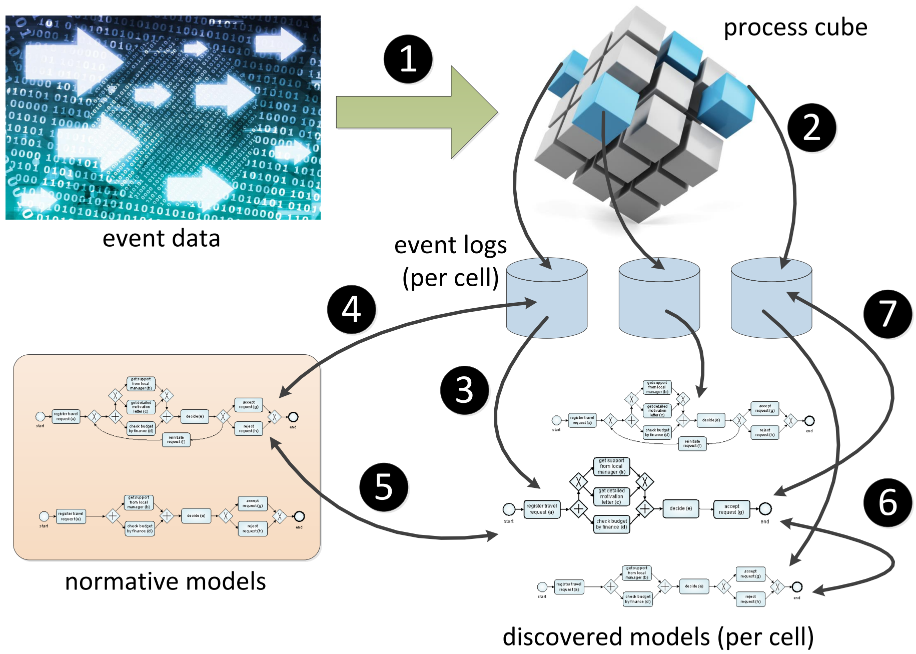

Fig. 1. Overview of comparative process mining using process cubes: ❶ store events in the process cube, ❷ materialize the events in a cell as an event log that can be analyzed, ❸ automatically discover models per cell (e.g., a BPMN or UML model), ❹ check conformance by replaying event data on normative (process) models, ❺ compare discovered models and normative models, ❻ compare discovered models corresponding to different cells, and ❼ compare different behaviors by replaying event data of one cell on another cell's model.

is driven by "raw" observed behavior instead of assumptions or aggregate data. Unlike classical data-driven approaches, it is truly process-oriented and relates events to high-level end-to-end process models [1].

Event logs serve as the starting point for process mining. An event log can be viewed as a multiset of *traces* [1]. Each trace describes the life-cycle of a particular *case* (i.e., a *process instance*) in terms of the *activities* executed. Often event logs store additional information about events, e.g., the *resource* (i.e., person or device) executing or initiating the activity, the *timestamp* of the event, or *data elements* recorded with the event.

2 Comparative Process Mining

Process mining can be used to analyze event data. The spectrum of available techniques is broad and includes techniques to automatically learn end-to-end process models, to check conformance, to analyze bottlenecks, to predict performance, etc. For an overview of available techniques see [1] or `processmining.org`. Here we would like to focus on *comparative process mining*, i.e., techniques that compare behavior either in the form of models or in the form of event logs [2, 5].

As Figure 1 shows, it all starts with event data. These event data are stored in a so-called *process cube* [2] with dimensions based on the event's attributes (see ❶ in Fig. 1). Note that in a process cube, there is no fixed assignment from events to cases (process instances). The same event may belong to different cells (e.g., people can work in two departments), different cases (e.g., a delivery may refer to multiple orders), and different processes (e.g., the sales and distribution processes may share common events). The dimensions may refer to groups of customers (gold versus silver customers), periods (2013 versus 2014), locations (Eindhoven versus Berlin), departments (sales versus procurement), performance (delayed or not), etc. These dimensions can be used to *slice*, *dice*, *roll-up*, and *drill-down* event data [2]. Events can be assigned to cases and standard attributes such as *activity*, *resource*, and *timestamp* can be chosen. Subsequently, cells can be materialized into concrete event logs (see ❷ in Fig. 1). Per cell different models can be *discovered* using dozens of different process mining techniques (see ❸ in Fig. 1). For example, one can automatically discover Petri nets or BPMN models from such event logs. Using *conformance checking* techniques one can also compare the event logs to normative process models (see ❹ in Fig. 1). These techniques quantify the conformance level and diagnose differences, e.g., highlighting activities that are skipped frequently [4]. It is also possible to compare discovered models with normative models (see ❺ in Fig. 1). Using the dimensions in the process cube one can also quickly compare different groups of cases, periods, locations, etc. For example, one can compare the models constructed for an array of cells (see ❻ in Fig. 1). What are the differences between the cases that got delayed and the cases that were able to meet the deadline? Why did the bottleneck shift from the back-office to the front-office in Spring 2014? Such questions can be answered using comparative process mining. Often we also compare a discovered process model for one cell with the event log of another cell (see ❼ in Fig. 1). Through conformance checking we can then analyze the differences at a very detailed level. For example, by replaying the event log of 2014 on the model constructed for 2013, we may see remarkable differences and immediately inspect the underlying event data.

3 Learning From Positive Deviants

As Figure 1 indicates, conformance checking can be done with respect to a normative model or the model constructed for another cell. The term "normative model" suggests that deviations are bad. However, there are many examples of *positive deviants*, i.e., cases that are non-conforming but also better performing (successful exceptions). The term "positive deviance" refers to approaches used to identify people (but also organizational entities and process variants) whose uncommon but successful behaviors or strategies enable them to find better solutions to a problem than their peers [6]. Positive deviance has been been applied in healthcare, education, agriculture, public administration, production, and services. The concept is simple: *look for outliers who succeed against all odds rather than sicking to a normative process model.* Comparative process mining –as explained using Fig. 1– is a powerful tool to distinguish between positive deviants, mainstream behavior, and negative deviants.

Process discovery and conformance checking techniques have matured over the last decade and are well-supported by the process mining framework *ProM* (`processmining.org`). However, better support for process cubes and an improved symbiosis between data and process mining are needed to provide a comprehensive toolbox for positive deviance. This way we can truly exploit the torrents of event data surrounding us.

References

[1] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* Springer-Verlag, Berlin, 2011.

[2] W.M.P. van der Aalst. Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. In M. Song, M. Wynn, and J. Liu, editors, *Asia Pacific Conference on Business Process Management (AP-BPM 2013)*, volume 159 of *Lecture Notes in Business Information Processing*, pages 1–22. Springer-Verlag, Berlin, 2013.

[3] W.M.P. van der Aalst. Data Scientist: The Engineer of the Future. In K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, editors, *Proceedings of the I-ESA Conference*, volume 7 of *Enterprise Interoperability*, pages 13–28. Springer-Verlag, Berlin, 2014.

[4] W.M.P. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying History on Process Models for Conformance Checking and Performance Analysis. *WIREs Data Mining and Knowledge Discovery*, 2(2):182–192, 2012.

[5] W.M.P. van der Aalst, S. Guo, and P. Gorissen. Comparative Process Mining in Education: An Approach Based on Process Cubes. In J.J. Lesage, J.M. Faure, J. Cury, and B. Lennartson, editors, *12th IFAC International Workshop on Discrete Event Systems (WODES 2014)*, IFAC Series, pages PL1.1–PL1.9. IEEE Computer Society, 2014.

[6] R. Pascale, J. Sternin, and M. Sternin. *The Power of Positive Deviance: How Unlikely Innovators Solve the World's Toughest Problems.* Harvard Business Review Press, 2010.