Business Process Mining for Collaborative Service-Oriented Systems – "Duality" of Process Representations and the Need for Statistical Treatment

Jörg Becker, Dominic Breuker

Department of Information Systems / ERCIS at the University of Muenster, Germany
Leonardo-Campus 3, 48149 Muenster, Germany
{becker,breuker}@ercis.uni-muenster.de

Abstract. Service-oriented systems are deployed by companies to support business processes, especially in inter-organizational collaborative settings. Process mining provides techniques to visualize and understand the emergent behavior in such systems based on data, as compared to what employees believe it is like. In highly unstructured settings however, these techniques have to deal with incomplete data, which still is a demanding challenge. In order to address it, we propose an alternative "dual" interpretation of mining results and outline the theoretical basis upon which process mining techniques could be extended and modified to deliver such results.

Keywords : Service-Orientation, Business Process Mining, Incompleteness

1 Motivation

With today's pressure to rapidly adapt to highly dynamic business environments, collaborations between different business partners have to be set up fast and in a flexible manner. Ad-hoc formation of virtual teams is an increasingly observable collaboration pattern. Technology can be considered one of the main facilitators of this development [1]. Service-oriented systems are perceived as a particularly suitable means of implementing collaborative services virtual teams rely on, including knowledge and resource sharing as well as communication and interaction [2].

For companies, it is important to be aware of their work practices in order to manage and align their activities. Business processes, codified within models, constitute a popular concept to do this [3]. Creating adequate business process models though is typically a laborious task as comprehensive interviews with process participants have to be conducted in order to identify and model the processes' structure [4]. Process mining techniques constitute a data-driven alternative. They allow analyzing business processes as they actually take place, provided that event data can be obtained from information systems involved in the processes' execution [5]. Consequently, process mining techniques could be applied to mine collaborative service behavior and to create business process models of what is going on in a company or even in virtual teams spanning multiple organizations.

Traditional modeling techniques such as Event-driven Process Chains (EPC) or the Business Process Model and Notation (BPMN) are successful in representing business processes of repetitive and well-standardized nature. Unstructured processes though are seen as hardly amenable to traditional process modeling due to uncertainty regarding their outcome as well as the steps and resources needed to produce it [6]. With process mining techniques being designed for processes of the first kind, the question is to which extent they are applicable to unstructured collaborative processes in service-oriented systems that belong more to the latter.

2 Related Work

Research investigating process mining to the field of services includes technical aspects such as logging in service-oriented architectures [7] but also case studies, e.g., about using process mining in the IBM WebSphere environment [8]. Challenges arising in this context include correlations between related process instances as well as restricted service behavior due to context [9]. Other challenges tackled in the literature include identifying events belonging to the same process instance [10].

However, these works target well-structured processes. With respect to collaborative service-oriented systems, little research has been done. Incompleteness, describing a situation in which a mining algorithm is provided with a dataset not including all necessary information, is a challenge process mining techniques must deal with. It is of particular importance in collaborative, unstructured settings [11]. The obvious reason is that the huge number of possibilities for performing an unstructured process leaves no hope for obtaining an event log enumerating them all. For this and other reasons, some researchers move away from mining holistic process descriptions towards aggregated features (e.g., number of interactions between individuals) [12].

3 Research Outline

The question we want to investigate is if it is possible to adapt process mining techniques in a way such that they can be applied in settings in which event logs are expected to be far from complete. Naturally, we cannot expect to generate an exact description of the underlying real-world process. This raises the question what any process model generated on an incomplete dataset is supposed to represent. A possible answer can be obtained through the following line of reasoning.

The normal way of thinking about processes models is that they specify the allowed behavior of a system. As an example, consider a simple process in which activity A is performed first, and then either activity B or C is performed after that, which both terminates the process. This *positive view* defines the process as a set of possibly observable instances: $\{AB, AC\}$. In such a setting, it makes sense to apply an algorithm to learn this allowed behavior. In a dual way though, one could equivalently think of a *negative view* that defines the process as the set of unobservable instances. In the example above, $\{AA, BA, BB, BC, CA, CB, CC, \dots\}$ would be this set. While this set will be huge for highly structured processes, the opposite might be the case for

unstructured ones. For this reason, algorithms searching for disallowed behavior in unstructured processes might use scarce data more efficiently.

But how can an algorithm infer disallowed behavior if only allowed behavior is observed in event logs? The answer is delivered by statistics. If direct negative evidence is unavailable one can work with indirect negative evidence instead, provided one is willing to assume a suitable statistical model. To illustrate this, consider the example of an unfair coin always showing heads when tossed. This could be inferred directly from the information that observing tails is the impossible event. Alternatively, one could observe that heads occurs surprisingly often, with each additional heads increasing the confidence that tails is a highly unlikely event. Statistical tests provide an established theoretical framework to incorporate such reasoning into process mining algorithms. Investigating how this can be accomplished is our approach to developing process mining techniques applicable to unstructured, collaborative service-oriented systems.

References

1. Lipnack, J., Stamps, J.: Virtual teams: People working across boundaries with technology. John Wiley & Sons, New York (2000)
2. Jerstad, I., Dustdar, S., Thanh, D. V.: A service oriented architecture framework for collaborative services. In: 14th IEEE International Workshops on Enabling Technologies Infrastructure for Collaborative Enterprise (WETICE'05). Linköping, Sweden (2005) 121–125
3. van der Aalst, W.M.P., Hofstede, A.H.M. ter, Weske, M.: Business Process Management: A Survey. Lecture Notes in Computer Science 2678 (2003) 1–12
4. Kettinger, W.J., Teng, J.T.C., Guha, S.: Business Process Change: A Study of Methodologies, Techniques, and Tools. MIS Quarterly 21(1) (1997) 55–80
5. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin / Heidelberg (2011)
6. Seidel, S., Müller-Wienbergen, F., Rosemann, M.: Pockets of Creativity in Business Processes. Communications of the Association for Information Systems 27(1) (2010) 415–436
7. Dustdar, S., Gombotz, R.: Discovering web service workflows using web services interaction mining. International Journal of Business Process Integration and Management 1(4) (2006) 256–266
8. van der Aalst, W.M.P., Verbeek, H.M.W.: Process Mining in Web Services: The WebSphere Case. IEEE Data Eng. Bull 31(3) (2008) 45–48
9. van der Aalst, W.M.P..: Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior. IEEE Transactions on Service Computing (2012)
10. Motahari-Nezhad, H.R., Saint-Paul, R., Casati, F., Benatallah, B.: Event correlation for process discovery from web service interaction logs. The VLDB Journal 20(3) (2011) 417–444
11. van der Aalst, W.M.P.: Exploring the CSCW spectrum using process mining. Advanced Engineering Informatics 21(2) (2007) 191–199
12. Truong, H.L., Dustdar, S.: Online Interaction Analysis Framework for Ad-Hoc Collaborative Processes in SOA-Based Environments. Transactions on Petri Nets and Other Models of Concurrency 2 (2009) 260–277