Comparing and Aligning Process Representations (Extended Thesis Abstract)

Han van der Aa*

Department of Computer Sciences, Vrije Universiteit Amsterdam, The Netherlands

Abstract. Having access to the right information on business processes is crucial to the proper and efficient execution of all sorts of activities, such as the assessment of mortgage applications, manufacturing of goods, as well as the treatment of patients. A major challenge here is that information related to a single process is often spread out over various models, documents, and systems. This fragmentation can have disastrous consequences for an organizations operations. It can, for example, lead to delays, wastes of money, and even violations of rules and laws. The work presented in this thesis tackles these problems with algorithms that can automatically compare process information stemming from various sources. These techniques, among others, enable the detection of contradictions between the sources and improve the ability of organizations to monitor their compliance to rules and regulations.

1 Motivation

Processes within organizations can be highly complex chains of inter-related steps, involving numerous stakeholders and information systems [6]. Due to this complexity, having access to the right information is vital to the proper execution and effective management of an organizations business processes [7]. A major threat to this information need is that the information on a single process, also referred to as *process information*, is often spread out over numerous models, documents, and systems. We shall refer to this condition as the *fragmentation of process information*.

This fragmentation occurs because organizations typically use multiple *informational artifacts* to provide process information to stakeholders with different informational needs or preferences [7, 9]. However, the use of multiple informational artifacts also poses considerable challenges to organizations. First, the fragmentation of process information can increase the effort required to access desired or necessary information. In particular, users may have to browse through numerous systems and documents to find the information they need [10, p.4]. Second, fragmentation can result in the provision of incorrect process information to users, for example because artifacts have become outdated. When users execute a process based on invalid information, they may perform the process in an incorrect manner. Such noncompliant acts can have severe consequences for organizations, including reduced productivity, a loss of control over processes, and even financial penalties imposed by authorities.

* Promotor: Prof. Dr. Ir. Hajo A. Reijers, Co-Promotor: Dr. Henrik Leopold

Despite the severity of these issues, there is only limited support for organizations to effectively deal with the negative effects of the fragmentation of process information. Furthermore, existing techniques that provide such necessary support only focus on specific situations, involving highly-structured process information in the form of process models and event logs. Process information in less-structured formats, such as natural language documents, is largely ignored by existing work. Consequently, there is a considerable gap between the way in which organizations capture information on their processes and the support that exists for this situation. As a result of this research gap, organizations still struggle with consequences caused by the fragmentation of process information.

2 Contributions

The main contributions of this thesis revolve around the definition of five techniques that focus on the comparison and alignment of process information in different informational artifacts. The techniques that we develop enable organizations to execute and maintain their process information more efficiently and to ensure that processes are executed according to their specification. Each technique addresses a specific scenario involving multiple informational artifacts in different representation formats:

2.1 Technique 1: Comparing Process Models to Textual Descriptions.

Motivation. Many organizations maintain textual process descriptions alongside graphical process models. Although this makes process information accessible to various stakeholders, there is a clear risk that model and text become misaligned when changes are not applied to both descriptions consistently. For organizations with hundreds of different processes, the effort required to manually identify and clear up such conflicts is considerable.
Solution. We address this problem with a technique that automatically identifies inconsistencies between a process model and a corresponding textual description. We developed tailored techniques to extract relevant process information from textual descriptions. Furthermore, our technique uses specific metrics, so-called *predictors*, to quantify the likelihood that the two representation formats contain inconsistencies.
Evaluation. A quantitative evaluation with a collection of 53 real-world processes shows that our technique can indeed successfully identify inconsistencies in model-text pairs. For instance, we observed that our technique identifies all model-text pairs containing missing activities with a precision of 0.70. Therefore, our proposed inconsistency-detection technique can be used by organizations to quickly identify inconsistencies in their process repositories. As a result, our technique supports organizations to maintain and improve the quality of their process documentation in an efficient manner.
Published as: [3]

2.2 Technique 2: Conformance Checking based on Uncertain Event-to-Activity Mappings

Motivation. Conformance checking enables organizations to automatically identify compliance violations based on the analysis of observed event data. A crucial require-

ment for conformance-checking techniques is that observed events can be mapped to normative process models used to specify allowed behavior. Without a mapping, it is not possible to determine if observed an event trace conforms to the specification or not. A considerable problem in this regard is that establishing a mapping between events and process model activities is an inherently uncertain task. Since the use of a particular mapping directly influences the conformance of an event trace to a specification, this uncertainty represents a major issue for conformance checking.
Solution. We developed a conformance-checking technique that avoids the problems associated with the need to select a single mapping. This is achieved by considering the entire spectrum of possible mappings generated by event-to-activity mapping techniques and capturing this spectrum in a behavioral space. Our probabilistic conformance-checking metric then provides insights into the fraction of conforming mappings, as well as useful diagnostic information.
Evaluation. A quantitative evaluation based on a collection of 557 real-world process models demonstrated that our technique can be used to obtain results in a vast number of cases where traditional conformance-checking techniques fail to do so. In particular, our technique was able to provide deterministic results for up to 50% more cases than traditional techniques.
Published as: [2]

2.3 Technique 3: Dealing with Ambiguity in Textual Process Descriptions.

Motivation. Textual process descriptions are widely used in organizations since they can be created and understood by virtually everyone. Because of their widespread use, they also provide a valuable source for process analysis, such as compliance checking. However, the inherent ambiguity of natural language impedes the automated analysis of textual process descriptions. While human readers can use their context knowledge to correctly understand statements with multiple possible interpretations, automated tools currently have to make assumptions about their correct meaning. As a result, conformance-checking techniques are prone to draw incorrect conclusions about the proper execution of a process.
Solution. To provide a comprehensive solution to these reasoning problems, we introduced the concept of a *behavioral space* as a means to deal with behavioral ambiguity in textual process descriptions. A behavioral space captures all possible interpretations of a textual process description in a systematic manner. Thus, it avoids the problem of focusing on a single, possibly incorrect interpretation.
Evaluation. We use a quantitative evaluation with a set of of 47 textual process descriptions to demonstrate that a behavioral space strikes a reasonable balance between ignoring ambiguous statements and imposing fixed interpretations on them. Furthermore, we demonstrated the usefulness of a semi-automated pruning technique to quickly reduce the level of uncertainty remaining in conformance-checking results.
Published as: [4]

2.4 Technique 4: Transforming and Aligning Process Performance Indicators.

Motivation. Monitoring process performance is an important means for organizations to identify opportunities to improve their operations. The definition of suitable Process

Performance Indicators (PPIs) is a crucial task in this regard. Because PPIs need to be in line with strategic business objectives, the formulation of PPIs is a managerial concern. Managers typically start out to provide relevant indicators in the form of natural language PPI descriptions. Therefore, considerable time and effort have to be invested to transform these descriptions into PPI definitions that can actually be monitored.
Solution. We developed a technique that transforms an unstructured natural language PPI description into a structured notation that is aligned with the implementation underlying a business process. To achieve this, the technique builds on hidden markov models as linguistic parsers to identify the parts of a PPI description that correspond to slots in a PPI template. Then, we use semantic similarity measures and semantic constraints to fill the slots with the appropriate values belonging to particular domains.
Evaluation. We evaluated the performance of our technique with a set of 129 real-world PPI descriptions from various industrial sources. The evaluation revealed that the structured PPI definitions generated by our technique are a good approximation of those created manually by experts. In particular, the technique achieves precision and recall values of, respectively, 0.89 and 0.82. Therefore, our technique represents a viable, automated alternative to an otherwise highly laborious and time-intensive, manual task. This enables organizations to more efficiently monitor the performance of their business processes and continuously adapt their monitoring activities to changing business needs.
Published as: [5]

2.5 Technique 5: Process Model Matching using Event-Log Information.

Motivation. Process model matching provides the basis for many process analysis techniques such as inconsistency detection and process querying. The matching task refers to the automatic identification of correspondences between activities in two process models. Despite the existence of a plethora of matching techniques, it has been shown that their results leave room for improvement [8]. A possible cause for this is that existing process model matching techniques focus exclusively on information related to the *specification* of processes, typically by just considering the information contained in process models themselves.
Solution. We present techniques that exploit event-log information for process model matching. *instance-based process matching*, which specifically focuses on information related to instances of a process. In particular, we introduce six similarity metrics that each use a different type of instance information stored in the event logs associated with processes. The matchers quantify similarity according to: event positions, frequencies, durations, attribute names, attribute values, and prerequisites. The proposed metrics can be used as standalone matching techniques or to complement existing process model matching techniques.
Evaluation. We demonstrated the usefulness of these similarity metrics through a quantitative evaluation using real-world data. The evaluation showed that by just considering the information specific to event logs, the introduced matchers can identify a considerable number of correspondences between event classes. This performance was particularly achieved when the six matchers were combined into a single matching ensemble.

As an illustration, the results show that users would need to evaluate just 4% out of all possible correspondences in order to identify nearly 70% of the actual correspondences.
Published as: [1]

3 Implications

The work presented in the thesis has a variety of implications for practice and research. For practice, the work has several implications for organizations that strive to improve their organizational efficiency and process quality. It achieves this by providing insights helpful for the improved management of process information, through technical solution for the efficient resolution of inconsistencies among process representation formats, and through improved detection of business process noncompliance.

For research, the thesis laid the groundwork for the comparison and alignment of process information in semi-structured and unstructured formats. Furthermore, it provides a foundation for process analysis and reasoning in the presence of behavioral uncertainty. Finally, it presents opportunities for the improvement of process model matching techniques through the consideration of event-log information.

As such, the presented work opens up possibilities for further research that pursues the efficient use and maintenance of process information within organizations through both technological and organizational measures.

References

1. Van der Aa, H., Gal, A., Leopold, H., Reijers, H.A., Sagi, T., Shraga, R.: Instance-based process matching using event-log information. In: International Conference on Advanced Information Systems Engineering. pp. 283–297. Springer (2017)
2. Van der Aa, H., Leopold, H., Reijers, H.A.: Checking process compliance on the basis of uncertain event-to-activity mappings. In: International Conference on Advanced Information Systems Engineering. pp. 79–93. Springer (2017)
3. Van der Aa, H., Leopold, H., Reijers, H.A.: Comparing textual descriptions to process models: The automatic detection of inconsistencies. Information Systems 64, 447–460 (2017)
4. Van der Aa, H., Leopold, H., Reijers, H.A.: Checking process compliance against natural language specifications using behavioral spaces. Information Systems (in press) (2018)
5. Van der Aa, H., Leopold, H., del Rio-Ortega, A., Resinas, M., Reijers, H.A.: Transforming unstructured natural language descriptions into measurable process performance indicators using hidden markov models. Information Systems 71, 27–39 (2017)
6. Van der Aalst, W.M.P., Ter Hofstede, A.H., Weske, M.: Business process management: A survey. In: International conference on business process management. pp. 1–12. Springer (2003)
7. Browning, T.R.: On the alignment of the purposes and views of process models in project management. Journal of Operations Management 28(4), 316–332 (2010)
8. Cayoglu, U., Dijkman, R.M., Dumas, M., Fettke, P., Garcıa-Banuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P., et al.: The process model matching contest 2013. In: 4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13) (2013)
9. Figl, K., Recker, J.: Exploring cognitive style and task-specific preferences for process representations. Requirements Engineering 21(1), 63–85 (2016)
10. Michelberger, B.: Process-oriented information logistics: aligning process information with business processes. Ph.D. thesis, Ulm University (2015)