Towards the Discovery of Object-Aware Processes

Marius Breitmayer and Manfred Reichert

Institute of Databases and Information Systems, Ulm University, Germany
{marius.breitmayer,manfred.reichert}@uni-ulm.de

Abstract. There has been a huge body of research in order to reduce manual efforts in creating executable process models through the automated discovery of process models from the event logs created by information systems. Regarding activity-centric processes, such event logs comprise case ids and events related to the execution of process activities. However, there exist alternative process management paradigms, such as object-aware processes, for which existing algorithms fail to discover a sound model. These algorithms do not treat data as first-class citizens, but solely rely on the information from event logs. In consequence, existing discovery algorithms are insufficient for discovering object-aware processes. To address this issue, discovery algorithms need to consider additional data sources (e.g., existing forms). This paper discusses the need for dedicated discovery techniques in object-aware processes.

Keywords: object-aware processes, process mining, process discovery

1 Introduction

Despite the many mining approaches that exist for activity-centric processes, adequate support for discovering data-centric process models, e.g., in the context of artifact-centric processes [7], case handling [2], or object-aware processes [8], is still lacking. While an activity-centric process model consists of a sequence of activities that need to be executed in a defined order, a data-driven and -centric process allows for greater flexibility through the use of declarative process rules and generated forms [15,6]. Current process discovery algorithms are able to discover the schema of an activity-centric process from an event log, whereas information about the internal logic of the activities (e.g., user forms or data required for an activity) is often neglected. As data is treated as first-class citizen in data-centric (e.g., object-aware) process management, the discovery of corresponding models should consider this issue as well.

To understand the nature of the problem at hand, a short introduction into data-centric and object-aware process management becomes necessary. PHILharmonicFlows, our approach to data-centric processes, introduces the concepts of *objects*, *object behavior,* and *object interaction.* For each business object present in a real-world business process, one such object exists. The latter comprises data, represented by *attributes,* and a state-based process model describing the

object behavior in terms of an *object lifecycle* model. When data becomes available during runtime, this enables transitions between the various states of the lifecycle process, i.e., execution is data-driven. In the e-learning system *PHoodle,* a practical application of the PHILharmonicFlows system [5], examples of business objects include *Submission*, *Exercise*, and *Lecture* (see Fig. 1 for the respective data model). In turn, when the values of certain attributes, such as *Points* or *Feedback,* become available at runtime, this enables the transition between the states of a lifecycle process (see Fig. 2). Finally, the interactions between object lifecycles are managed by coordination processes [14].
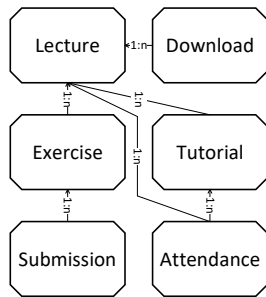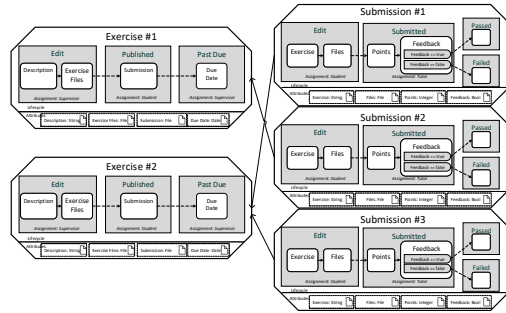

Fig. 1. Data Model

Fig. 2. Objects with Lifecycles and Interaction

2 Related Work

Process discovery summarizes techniques that leverage information from event logs to discover process models [3]. For activity-centric processes, there exist a variety of approaches (see [1,17] for an overview). Various algorithms use event logs as input to discover an activity-centric process model. Regarding data-centric processes [16], however, there only exist few approaches for discovering process models. [12] describes an approach for discovering artifacts and their lifecycles from structured datasets as opposed to lifecycle-enabled objects in our approach. In turn, [9] deals with methods for discovery of artifacts and the interactions between them; additionally, an evaluation based on real-life datasets from ERP systems is provided. In turn, [13] decomposes the problem of artifact lifecycle discovery such that existing process mining algorithms can be applied. The construction of data and object models from different data structures (e.g., databases, legacy systems) has been investigated in reverse engineering [4,10]. While database reverse engineering reconstructs logical or conceptual models, other aspects of data-driven process management are neglected (e.g., lifecycles or the interactions between object lifecycles). An approach to automatically generate event logs from databases is described in [11]. Since data is treated as a first-class citizen in object-oriented process management, additional information (i.e., data sources) need to be considered to discover an object-aware process.

3 Research direction

In our PHILharmonicFlows framework, an object-aware process consists of a *data model*, one *lifecycle model* for each object, and a *coordination process* enforcing constraints regarding object interactions [8]. In order to discover an executable object-aware process, all three aspects need to be considered. For the discovery of various aspects of object-aware processes (e.g., relations between objects or states of a lifecycle), solely considering event logs is not sufficient and, hence, additional data sources need to be taken into account. For example, the data model underlying an object-aware process provides the foundation for both lifecycles and object coordination [5].

The **first step** during process discovery is to identify objects, including their attributes and relations. Note that the structure of a normalized relational database, to a certain degree, is comparable to a data model, which offers the opportunity to discover the data model from the structure (i.e., the create table statements) of a database. Each table in the database may, but does not have to correspond to an object in the data model, whereas columns of a table represent the attributes of an object. One-to-many relations between tables can be used to identify relations between the objects of a data model. Additionally, relations can be used as an indicator if a table corresponds to a correct object.

After discovering the data model, the object lifecycles need to be discovered in the **second step**. Based on the attributes from the data model, lifecycle discovery shall deliver object states as well as the transitions between them. In general, a lifecycle process may enter another state, if all necessary data (i.e., attribute values) are available. In particular, lifecycle states cannot be discovered from event logs, whose entries solely refer to activities due to the mismatch between states (i.e., defined by attributes) and cases (i.e., a collection of activities). To tackle this mismatch, discovery algorithms for object lifecycles, suitable event log preprocessing (e.g., splitting an event log into event logs for each object), and additional data sources (e.g., forms of existing information systems) need to be considered as well during the discovery process.

The **third step** in discovering an object-aware process is to unravel the coordination of interactions between objects (e.g., a submission may only be created if the corresponding exercise is in state published). As object interaction can only be discovered with the data model and the lifecycles present, their discovery is a secondary problem for now.

4 Conclusion

This paper discusses the need for spending research efforts on the discovery of object-aware process models. As major advantage, the discovery of an object-aware processes allows to identify the underlying logic of a process. Finally, due to the strong linkage between process and data in object-aware processes, it is possible that not every aspect of each element (i.e., data model, lifecycles, and coordination) may be discovered from the presented data sources and, therefore further research is of utmost importance.

References

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, 2 edn. (2016)
2. van der Aalst, W.M.P., Weske, M., Grünbauer, D.: Case handling: a new paradigm for business process support. DKE **53**(2), 129–162 (2005)
3. van der Aalst, W.M.P., et al.: Process mining manifesto. In: Int'l Conf on BPM'11. pp. 169–194 (2011)
4. Alhajj, R.: Extracting the extended entity-relationship model from a legacy relational database. Information Systems **28**(6), 597 – 618 (2003)
5. Andrews, K., Steinau, S., Reichert, M.: Engineering a highly scalable object-aware process management engine using distributed microservices. In: Int'l Conf on CoopIS'18. pp. 80–97 (2018)
6. Andrews, K., Steinau, S., Reichert, M.: Enabling runtime flexibility in data-centric and data-driven process execution engines. Information Systems p. 101447 (2019)
7. Cohn, D., Hull, R.: Business artifacts: A data-centric approach to modeling business operations and processes. IEEE Data Eng. Bull. **32**(3), 3–9 (2009)
8. Künzle, V., Reichert, M.: PHILharmonicFlows: towards a framework for object-aware process management. J of Soft Maint & Evo **23**(4), 205–244 (2011)
9. Lu, X., Nagelkerke, M., van de Wiel, D., Fahland, D.: Discovering interacting artifacts from ERP systems. IEEE Trans Serv Com **8**(6), 861–873 (2015)
10. Mfourga, N.: Extracting entity-relationship schemas from relational databases: a form-driven approach. In: WCRE'97. pp. 184–193 (1997)
11. de Murillas, et al.: Case notion discovery and recommendation: automated event log building on databases. Know & Inf Sys (2019)
12. Nooijen, E.H.J., van Dongen, B.F., Fahland, D.: Automatic discovery of data-centric and artifact-centric processes. In: Int'l Conf on BPM'12. pp. 316–327 (2012)
13. Popova, V., Fahland, D., Dumas, M.: Artifact lifecycle discovery. Int'l J of Coop Inf Sys **24**(01), 1550001 (2013)
14. Steinau, S., Andrews, K., Reichert, M.: Modeling process interactions with coordination processes. In: CoopIS'18. pp. 21–39. LNCS, Springer (2018)
15. Steinau, S., Andrews, K., Reichert, M.: Executing lifecycle processes in object-aware process management. In: Data-Driven Process Discovery and Analysis. pp. 25–44. Springer (2019)
16. Steinau, S., Marrella, A., Andrews, K., Leotta, F., Mecella, M., Reichert, M.: DALEC: A framework for the systematic evaluation of data-centric approaches to process management software. Softw & Sys Modeling **18**(4), 2679–2716 (2019)
17. Weerdt, J.D., Backer, M.D., Vanthienen, J., Baesens, B.: A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. Inf Sys **37**(7), 654 – 676 (2012)