Sub-Model Freezing During Incremental Process Discovery in Cortado (Extended Abstract)

Daniel Schuster
Fraunhofer Institute for Applied Information Technology FIT
Sankt Augustin, Germany
daniel.schuster@fit.fraunhofer.de
RWTH Aachen University
Aachen, Germany

Sebastiaan J. van Zelst
Fraunhofer Institute for Applied Information Technology FIT
Sankt Augustin, Germany
sebastiaan.van.zelst@fit.fraunhofer.de
RWTH Aachen University
Aachen, Germany

Wil M. P. van der Aalst
Fraunhofer Institute for Applied Information Technology FIT
Sankt Augustin, Germany
RWTH Aachen University
Aachen, Germany
wvdaalst@pads.rwth-aachen.de

***Abstract*—Conventional process discovery algorithms are fully automated and work as a black box from the user's perspective. Event data is fed into the discovery algorithm, and a process model is returned. Interactive process discovery is about breaking this black-box approach of conventional process discovery and involving the user during the discovery, i.e., adopting the principles of hybrid intelligence in process discovery. The central idea is to exploit the user's knowledge of the process to be discovered within the discovery phase to obtain better models. The software tool *Cortado* allows for the incremental discovery of a process model based on user-selected process behavior. In this paper, we present the implementation of *sub-model freezing*, i.e., a novel form of user interaction during incremental process discovery, in Cortado.**

***Index Terms*—process mining, interactive process discovery, process models, hybrid intelligence**

I. Introduction

Process discovery, a key discipline of process mining [1], comprises algorithms that (automatically) learn a process model from event data. Since event data often have quality issues and are incomplete, i.e., only a fragment of the possible process behavior is captured, conventional process discovery algorithms often yield low-quality process models. To address these challenges, the field of interactive process discovery has emerged. The key idea is to utilize domain knowledge about the process to be discovered, in addition to the available event data, to discover process models of superior quality.

In [2], we introduced the first version of the software tool *Cortado*. Following an incremental process discovery approach, Cortado enables the user to gradually discover a process model from user-selected process behavior, i.e., event data. This incremental approach to process discovery allows the user to influence the discovery of a process model interactively. For a detailed description of Cortado's functionality, we refer to [2]. In recent work [3], we presented a novel form of user interaction in the context of incremental process discovery: *sub-model freezing*. In this paper, we present the realization of sub-model freezing within Cortado[1].

[1]Sub-model freezing is available from version 1.3.0, downloadable from https://cortado.fit.fraunhofer.de/

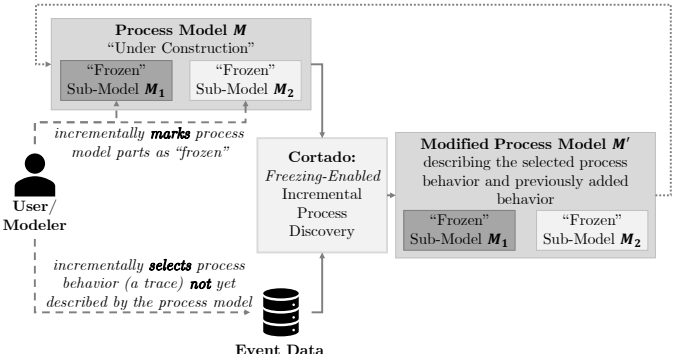
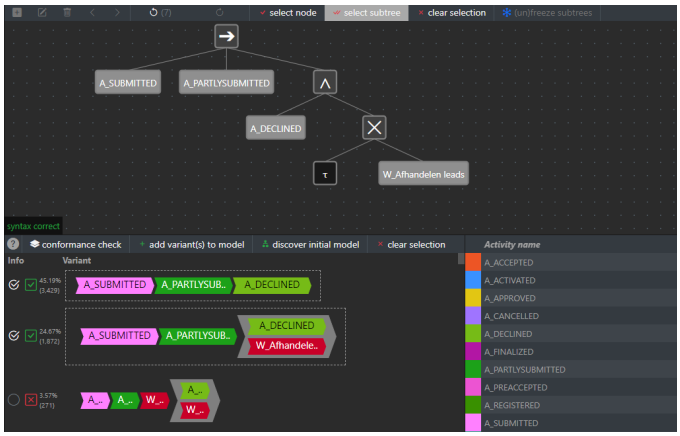


Fig. 1. Conceptual idea of sub-model freezing during incremental process discovery. Figure adapted from [3].

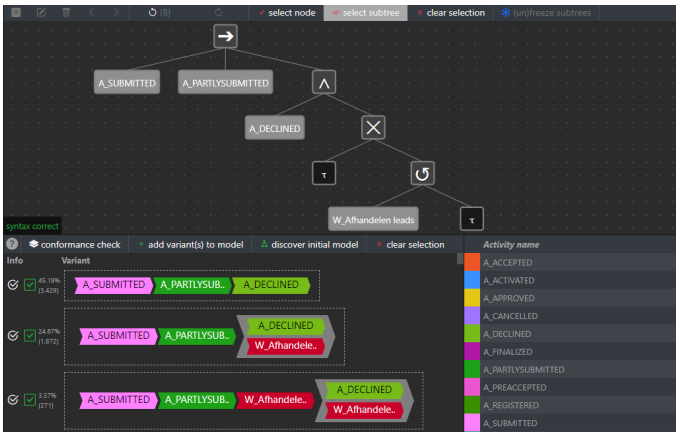
II. Sub-Model Freezing

In this section, we first outline the concept of sub-model freezing within incremental process discovery. Afterwards, we focus on the implementation of said technique in Cortado.

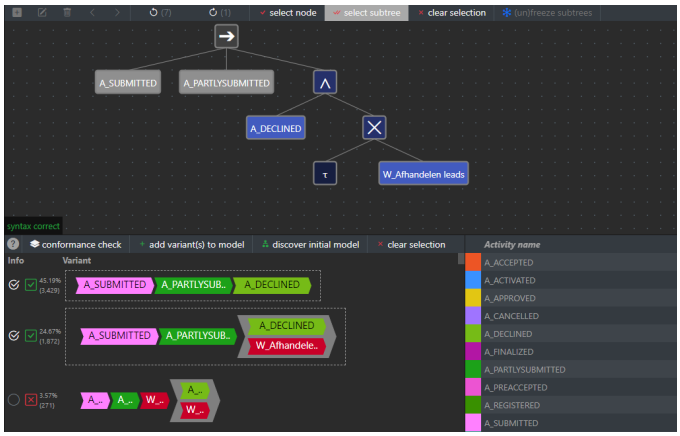
The theoretical foundations of sub-model freezing are introduced in [3]. Figure 1 visualizes the conceptual idea. Starting from an event log and an initial model M, which can also be discovered by Cortado, a user incrementally selects process behavior, i.e., trace variants, that are not yet described by the process model M. Additionally, the user has the option to freeze sub-models of M. For example, as indicated in Figure 1, the user freezes two sub-models, i.e., M_1 and M_2, of M. By freezing sub-models of M, the freezing-enabled incremental discovery approach implemented in Cortado ensures that the incrementally discovered process model M' contains M_1 and M_2. Without marking M_1 and M_2 as frozen, there is no guarantee that these sub-models will be present in the new model M' in identical form. Note that the incrementally discovered process model M' describes the selected trace variant plus previously selected trace variants. After one iteration, the user can incrementally add further trace variants to the model under construction. Note that the incrementally discovered model M' is used as an input in the next iteration, visualized by the dotted arc from M' to M in Figure 1. Further, the user can change



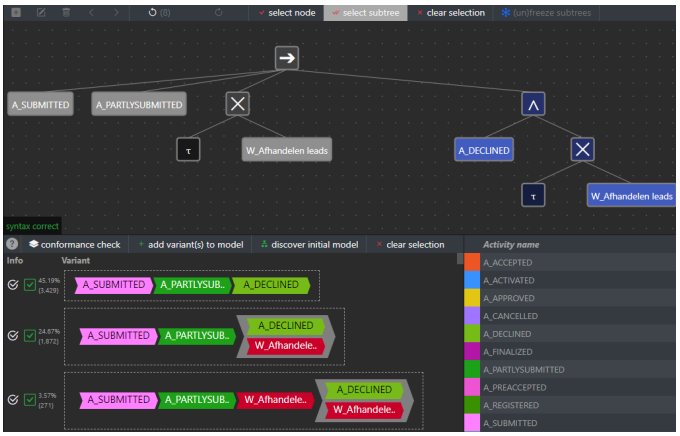
(a) *Without freezing:* Initial model that describes the first two variants (indicated by green check-marks, which are located left to each variant)



(b) *Without freezing:* Process model after the third variant has been incrementally added to the model shown in Figure 2a



(c) *With freezing:* Initial model as shown in Figure 2a with frozen process model part (frozen subtree is marked blue)



(d) *With freezing:* Process model after the third variant has been incrementally added to the model shown in Figure 2c.

Fig. 2. Example of an incremental process discovery step, i.e., adding a non-fitting trace variant to a model under construction, with/without freezing.

which sub-models are frozen in each iteration.

In Figure 2, we present screenshots of Cortado that are demonstrating the described incremental process discovery approach once without freezing (Figure 2a and 2b) and once with freezing (Figure 2c and 2d). In both cases, we use the same event data and the same initial model that describes the first two variants from the variant explorer, cf. Figure 2a and 2c. Note that Cortado uses process trees as a process model formalism. We refer to [4] for an introduction to process trees.

Figure 2b shows the process tree after adding the third variant from the variant explorer to the initial process tree. We observe that the algorithm added a loop on the activity `W_Afhandelen leads`. In Figure 2c, we see the same initial process tree where the user marked a subtree as frozen, highlighted in blue colors. After adding the the third variant to the initial process tree with frozen subtree, we observe that the resulting process tree is different compared to the one obtained without freezing. This time, the algorithm added an optional activity labeled with `W_Afhandelen leads` before the frozen subtree is executed. Note that the frozen subtree has not been altered by the algorithm, compared to the execution without freezing (Figure 2b). Further, note that both discovered process trees, i.e., with/without freezing, describe the three selected trace variants shown; however, they structurally differ.

III. Conclusion

In this paper, we presented the realization of sub-model freezing—a novel form of user interaction within incremental process discovery—in Cortado. Further, we highlighted the difference between using and not using the freezing option with an example.

References

[1] W. M. P. van der Aalst, *Process Mining - Data Science in Action, Second Edition*. Springer, 2016.

[2] D. Schuster, S. J. van Zelst, and W. M. P. van der Aalst, "Cortado—an interactive tool for data-driven process discovery and modeling," in *Application and Theory of Petri Nets and Concurrency*, ser. Lecture Notes in Computer Science, vol. 12734. Springer, 2021.

[3] ——, "Freezing sub-models during incremental process discovery," in *Conceptual Modeling*, ser. Lecture Notes in Computer Science, vol. 13011. Springer, 2021.

[4] ——, "Incremental discovery of hierarchical process models," in *Research Challenges in Information Science*, ser. Lecture Notes in Business Information Processing, vol. 385. Springer, 2020.