
In Log and Model We Trust? (extended abstract)

Andreas Rogge-Solti¹, Arik Senderovich², Matthias Weidlich³, Jan Mendling⁴ and Avigdor Gal⁵

Abstract: While models and event logs are readily available in modern organizations, their quality can seldom be trusted. Raw event recordings are often noisy, incomplete, and contain erroneous recordings. The quality of process models, both conceptual and data-driven, heavily depends on the inputs and parameters that shape these models, such as domain expertise of the modelers and the quality of execution data. The mentioned quality issues are specifically a challenge for conformance checking. Conformance checking is the process mining task that aims at coping with low model or log quality by comparing the model against the corresponding log, or vice versa. The prevalent assumption in the literature is that at least one of the two can be fully trusted. In this work, we propose a *generalized conformance checking* framework that caters for the common case, when one does neither fully trust the log nor the model. In our experiments we show that our proposed framework balances the trust in model and log as a generalization of state-of-the-art conformance checking techniques.

Keywords: process mining, conformance checking, model repair, log repair

1 Introduction

Business process management plays an important role in modern organizations that aim at improving the effectiveness and efficiency of their processes. To assist in reaching this goal, the research area of process mining offers multitude of techniques to analyze event logs that carry data from business processes. Process mining investigates the interplay among reality (*system*), its reported observations (*event log*), and a corresponding *process model* [BvDvdA14]. While reality is typically unknown, we are left with the need to reconcile the event log and the process model, where evidence of a certain behavior may only be present in one but not in the other.

Current conformance checking techniques are not capable of defining levels of trust for model and log to cater for uncertainty. Therefore, we consider the problem of optimally reconciling an event log with a process model, given an input event log and a model (if such exist) and our degree of trust in each. In this extended abstract we present the results of our work in [Ro16], where we outline that various process mining tasks can actually be regarded as special cases of this generic problem formulation. The problem formulation goes beyond locating misalignments between a process model and an event log by

¹ Wirtschaftsuniversität Wien, Vienna, Austria, firstname.lastname@wu.ac.at

² Technion–Israel Institute of Technology, Haifa, Israel sariks@tx.technion.ac.il

³ Humboldt University of Berlin, Germany firstname.lastname@hu-berlin.de

⁴ Wirtschaftsuniversität Wien, Vienna, Austria, firstname.lastname@wu.ac.at

⁵ Technion–Israel Institute of Technology, Haifa, Israel avigal@ie.technion.ac.il

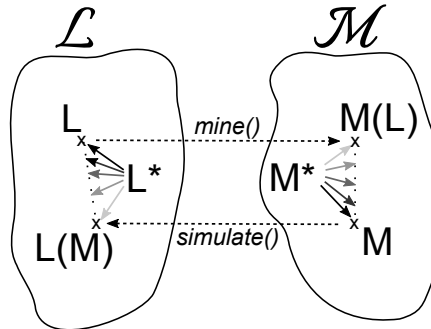


Figure 1: Conceptual sketch of the problem setting. [Ro16]

providing explanations of misalignments and categorizing them as one of a) anomalies in an event log, b) modeling errors, and c) unresolvable inconsistencies. This generalized conformance checking problem can be seen as the unification of conformance checking, model repair, and anomaly detection.

2 The Generalized Conformance Checking Problem

We consider the setting that next to the given input event log L and input process model M , we also have a trust level for the model π_M and a trust level for the log π_L . Latter reflect for instance the trust in the event recording mechanism, or the trust in the abilities of the modeler. In this setting, we are interested in finding the optimal (repaired) log L^* and (repaired) model M^* pair that best fit the input log L and model M , and also fit each other best according a distance measure (e.g., replay fitness).

To solve this problem, we propose a two-step divide-and-conquer approach. The main idea of this approach is to avoid the inherent complexity induced by the freedom to change the model or the log by sequentialization: first identifying changes in the model, before turning to changes applied to the log.

Our approach is outlined in Figure 1. We lift the problem into the model space by mining a model $M(L)$ by representing event logs as their discovered counterparts. Then, we approximate M^* by applying a greedy heuristic search in the space between the input model M and the mined model $M(L)$. If we strongly trust our input model, then, we respect that by not allowing to move too far away from the input model. If we do not trust our input model at all, we would end up with the mined model $M(L)$.

After approximating the optimal model M^* , we align the input log L to it using techniques like [vdAAvD12], and see which deviations remain. These misalignments then need to be classified into errors in the log, and remaining non-conformance between event log and model. Here, we use the trust in the log to determine the the share that will be corrected in the log. For example, if we trust our log to be correct entirely (trustlevel 1), we do not

Process mining task	Log Trust	Model Trust
Classical Process Discovery finds a model that best fits to the entire event log, e.g., the alpha algorithm [vdAWM04].	$\pi_L = 1$	$\pi_M = 0$
Heuristic Process Discovery algorithms apply preprocessing to the event log by discarding infrequent patterns [GvdA07, WvdADM06].	$0 < \pi_L < 1$	$\pi_M = 0$
Model Repair fixes deficient models due to e.g., a change in the system that is reflected in the log. For example [FvdA15].	$\pi_L = 1$	$0 < \pi_M < 1$
Conformance Checking. This task tries to find misalignments between event log and model. Example works include [RvdA08, vdAAvD12, Se16].	$\pi_L = 1$	$\pi_M = 1$
Log Repair. Given a trusted model and a noisy log, we modify the log until it conforms to the model [Ro13, RSK14, Wa15].	$0 < \pi_L < 1$	$\pi_M = 1$
“Happy Path” Simulation is complementary to heuristic process discovery. It is a theoretical use case where we do not trust infrequent parts of the model [MSS15].	$\pi_L = 0$	$0 < \pi_M < 1$
Process Simulation is complementary to process discovery, where we are given an untrustworthy empty log and a fully trustworthy model.	$\pi_L = 0$	$\pi_M = 1$
Garbage In, Garbage Out. When both the model and the log are untrustworthy, the best log and model tuple that fits them is any pair of model and log that fits each other, including an empty log and an empty model.	$\pi_L = 0$	$\pi_M = 0$
Generalized Conformance Checking is the focus of this paper. Instead of only detecting the misalignments, as in conformance checking, we also provide, where the model would best be adopted, and where the log would best be adopted for a better overall fit.	$0 < \pi_L$	$0 < \pi_M$

Table 1: Some process mining tasks cast as problem instances. [Ro16]

repair the log at all. If our trust in the input log is less (because it is based on noisy sensors for instance), then we repair a corresponding share of the misalignments in the event log.

Table 1 characterizes different areas of process mining and business intelligence that this generalized framework covers for different trust levels.

3 Conclusion

In this work, we presented a generalization of the conformance checking problem. It strives for a balance between two independent input parameters: the trust in the log quality, and the trust in the model quality. Specifically, when presented with an event log and a process model, generalized conformance checking attempts at repairing both according to

the initial trust levels, and returns an improved log-model pair. Generalized conformance checking is comparable to state-of-the-art model repair techniques in model quality measures. The full formalisation, evaluation results, and further details can be found in the original conference paper [Ro16].

References

- [BvDvdA14] Buijs, Joos C. A. M.; van Dongen, Boudewijn F.; van der Aalst, Wil M. P.: Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *International Journal of Cooperative Information Systems*, 23(01):1440001, 2014.
- [FvdA15] Fahland, Dirk; van der Aalst, Wil M.P.: Model repair – aligning process models to reality. *Information Systems*, 47:220 – 243, 2015.
- [GvdA07] Günther, Christian W.; van der Aalst, Wil M. P.: Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In: *Business Process Management*, pp. 328–343. Springer, 2007.
- [MSS15] Marquard, Morten; Shahzad, Muhammad; Slaats, Tijs: Web-based modelling and collaborative simulation of declarative processes. In: *Business Process Management*, pp. 209–225. Springer, 2015.
- [Ro13] Rogge-Solti, Andreas; Mans, Ronny S.; van der Aalst, Wil M. P.; Weske, Mathias: Improving documentation by repairing event logs. In: *The Practice of Enterprise Modeling*, pp. 129–144. Springer, 2013.
- [Ro16] Rogge-Solti, Andreas; Senderovich, Arik; Weidlich, Matthias; Mendling, Jan; Gal, Avigdor: In Log and Model We Trust? - A Generalized Conformance Checking Framework. In: *Business Process Management (BPM'16)*. 2016. (to appear).
- [RSK14] Rogge-Solti, Andreas; Kasneci, Gjergji: Temporal anomaly detection in business processes. In: *Business Process Management*, pp. 234–249. Springer, 2014.
- [RvdA08] Rozinat, Anne; van der Aalst, Wil M. P.: Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64 – 95, 2008.
- [Se16] Senderovich, Arik; Weidlich, Matthias; Yedidsion, Liron; Gal, Avigdor; Mandelbaum, Avishai; Kadish, Sarah; Bunnell, Craig A.: Conformance checking and performance improvement in scheduled processes: A queueing-network perspective. *Information Systems*, Forthcoming, 2016.
- [vdAAvD12] van der Aalst, Wil M. P.; Adriansyah, Arya; van Dongen, Boudewijn F.: Replaying History on Process Models for Conformance Checking and Performance Analysis. *WIREs: Data Mining and Knowledge Discovery*, 2:182–192, 2012.
- [vdAWM04] van der Aalst, Wil M. P.; Weijters, Ton; Maruster, Laura: Workflow mining: Discovering process models from event logs. *Knowledge and Data Engineering, IEEE Transactions on*, 16(9):1128–1142, 2004.
- [Wa15] Wang, Jianmin; Song, Shaoxu; Lin, Xuemin; Zhu, Xiaochen; Pei, Jian: Cleaning structured event logs: A graph repair approach. In: *Data Engineering (ICDE'15)*. IEEE, pp. 30–41, 2015.
- [WvdADM06] Weijters, A. J. M. M.; van der Aalst, Wil M. P.; De Medeiros, A. K. Alves: Process mining with the heuristics miner-algorithm. Technical Report 166, 2006.