Partial Order Resolution of Event Logs for Process Conformance Checking (Extended abstract)[1]

Han van der Aa[2], Henrik Leopold[3,4], Mathias Weidlich[5]

Context. Non-conformance in process execution poses a considerable threat to organizations. Conformance issues occur when the actual behavior of a process deviates from the desired one. Such differences can occur because information systems typically *support* process execution, but do not *enforce* a particular process. Rather, human interaction drives a process, giving people a certain flexibility in the execution of a particular case. To detect conformance issues, techniques for *conformance checking* have been introduced [Ca18]. They strive for the automatic detection of deviations between the observed and desired process behavior by comparing recorded event data to a normative process model.

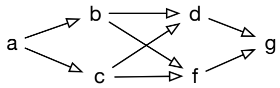
Problem. A key assumption of most conformance checking techniques is that all events of a case have timestamps that allow to infer a *total* order. However, factors like synchronization issues, manual recording of events, and unreliable data sensing, mean that this assumption is often violated in practice. E.g., in healthcare processes, only the date of certain treatments may be known, but not their specific point in time [Ma13]. As a result of this, events are only partially ordered, meaning that a single trace could actually reflect various possible total orders, i.e., different possible resolutions, as depicted in Figure 1. Since different resolutions can have different conformance results, the conformance of such traces becomes uncertain.

Contribution. Our work aims to still provide conformance results in the presence of such order uncertainty. Its idea is to use information from the entire event log to estimate the probability of each possible resolution of an uncertain trace. For this estimation, we argue that the probability of a specific resolution may be based on order information derived from similar traces contained in the event log. Intuitively, if a possible resolution denotes an execution order that is frequently observed for other traces, this resolution is expected to be more likely than one that denotes a less common sequence. By incorporating these probabilities in our approach, conformance checking is grounded in a stochastic model.

Behavioral models. Estimating resolution probabilities requires a careful selection of the abstraction level based on which traces are compared. This is because, depending on the complexity of a process and the degree of order uncertainty, it may be impossible to

[1] The original article has been published in Decision Support Systems [vdAa20b]

[2] University of Mannheim, Germany. han@informatik.uni-mannheim.de
[3] Kühne Logistics University, Hamburg, Germany. henrik.leopold@the-klu.org
[4] Hasso Plattner Institute, University of Potsdam, Germany
[5] Humboldt-Universität zu Berlin, Germany. matthias.weidlich@hu-berlin.de



$\pi_1 = \langle a, b, c, d, f, g \rangle$
$\pi_2 = \langle a, c, b, d, f, g \rangle$
$\pi_3 = \langle a, b, c, f, d, g \rangle$
$\pi_4 = \langle a, c, b, f, d, g \rangle$

Fig. 1: Partial order and trace resolutions for an uncertain trace $\langle \{a\}, \{b, c\}, \{d, f\}, \{g\} \rangle$.

observe the exact same sequence of activity executions in another trace, unaffected by order uncertainty. We cope with this issue by defining behavioral models that realize different levels of abstraction for the comparison of traces. Particularly, we propose models that determine the likelihood of a resolution based on: (1) fully identical traces, (2) identical sub-traces of length N, and (3) weak-order relations. These models differ in the notion of behavioral regularity that is used for the partial order resolution and, importantly, in their ability to derive information from traces that are also only partially ordered.

Evaluation results. Extensive evaluation experiments, using both real-world and synthetic data, reveal that a model focusing on sub-traces of length 2, i.e., 2-grams, yields the most accurate conformance checking results. When compared to existing baselines, we observe that our approach is more accurate, reducing the average error by 59%. Furthermore, we demonstrate that our approximation method, introduced to cope with the high computational complexity of conformance checking under uncertainty, leads to considerable gains in terms of efficiency, while having a negligible impact on the result accuracy.

Outlook. The discussed work is part of our ongoing research stream of conformance checking in the presence of uncertainty, in which we target three major sources of uncertainty: uncertainty (1) in the process specification [vdAa18], (2) in the mapping from specification to event data [vdAa20], and (3) now also uncertainty in the data itself. For all these directions, we see open research questions related to the perception and quantification of probabilistic results in conformance checking. For the specific direction presented here, we furthermore intend to explore additional behavioral models, involving data beyond the control flow, as well as techniques that use neural networks to learn how order uncertainty shall be resolved.

Bibliography

[Ca18] Carmona, Josep; van Dongen, Boudewijn F.; Solti, Andreas; Weidlich, Matthias: Conformance Checking - Relating Processes and Models. Springer, 2018.

[Ma13] Mans, Ronny S; Van der Aalst, Wil M P; Vanwersch, Rob JB; Moleman, Arnold J: Process mining in healthcare: Data challenges when answering frequently posed questions. In: ProHealth, pp. 140–153. Springer, 2013.

[vdAa18] Van der Aa, Han; Leopold, Henrik; Reijers, Hajo A: Checking process compliance against natural language specifications using behavioral spaces. Inf. Syst., 78:83–95, 2018.

[vdAa20] Van der Aa, Han; Leopold, Henrik; Reijers, Hajo A.: Efficient Process Conformance Checking on the Basis of Uncertain Event-to-Activity Mappings. TKDE, 32(5):927–940, 2020.

[vdAa20b] Van der Aa, Han; Leopold, Henrik; Weidlich, Matthias: Partial order resolution of event logs for process conformance checking. Decision Support Systems, 136:113347, 2020.