

A GENERAL FRAMEWORK FOR PRECISION CHECKING

JORGE MUNOZ-GAMA AND JOSEP CARMONA

Department of Software
Universitat Politècnica de Catalunya
C. Jordi Girona, 1. 08034, Barcelona, Spain
{jmunoz; jcarmona}@lsi.upc.edu

Received March 2011; revised October 2011

ABSTRACT. *Process conformance is the domain within Process Mining that addresses the adequacy between a model and a log of a system. It has four different dimensions: fitness, precision, generalization and structure. This paper presents a metric to evaluate the precision dimension: the extra behavior a formal model tolerates when confronted to a log. Additionally, two important factors are presented together with the metric value: confidence (an estimation of the stability of the metric value for a future window), and a severity assessment to the imprecisions detected. Several techniques are described to accomplish this, including a log-guided traversal of the model, the optimization of binary linear programming models to estimate the confidence, and multi-factored methods to determine the severity. The approach is implemented within an open-source Process Mining platform, and experimental results certify both the significance of the approach and the usefulness of the new factors proposed.*

Keywords: Process mining, Conformance checking, Business process management

1. **Introduction.** Process mining is a discipline that combines formal techniques that address the process perspective of a system, and data mining [1]. It has been an emerging area in the last decade. The reason for the success of process mining is simple: there is a lack of automation and formal analysis in the way nowadays information systems are designed and maintained. We live in a world where there is an exponential growth of data, and where software engineers must necessarily adapt to an always evolving situation. Process mining techniques are based on extracting information from *logs*, where a log is a set of traces observed in the execution of a system. These techniques address the process perspective of an information system in the following scopes: *process discovery* (find a formal model to represent a log), *process conformance* (evaluate the adequacy of a formal model in representing a log) and *process extension* (incorporate enhancements into the model). The work presented in this paper presents techniques for process conformance.

Process conformance comprises four dimensions [2]. The first dimension is called *fitness*, and addresses how much of the log is captured by the process model. The *precision* dimension instead prioritizes these models that represent as closely as possible the log. In contrast, the *generalization* dimension prioritizes models having some degree of abstraction in representing the log. Finally, the *structure* dimension refers to models minimal in structure which reflect the log.

In this work we propose techniques for estimating the precision of a formal model in representing a log. Importantly, the approach presented in this paper estimates the effort needed to achieve a better model (more precise), and identifies the mismatches between log and model. This may speed-up the modeling stage by iterative evaluation and enhancement of the processes within a system. Additionally, we incorporate discussions on how to adapt the techniques to models that are popular in *Process-Aware Information*