

Workshop Proceedings



Workshop on
**Algorithms & Theories for the
Analysis of Event Data (ATAED'2020)**

virtual workshop, June 24, 2020

supported by the IEEE Task Force on Process Mining

Satellite event of the conference

**41st International Conference on Application and Theory
of Petri Nets and Concurrency (Petri Nets 2020)**

Edited by
Wil van der Aalst, Robin Bergenthum, and Josep Carmona

Copyright © 2020 for the individual papers is held by the papers' authors.
Copying is permitted only for private and academic purposes.
This volume is published and copyrighted by its editors.

Preface

Ehrenfeucht and Rozenberg defined regions more than 30 years ago as sets of nodes of a finite transition system. Every region relates to potential conditions that enable or disable transition occurrences in an associated elementary net system. Later, similar concepts were used to define regions for Petri nets from languages as well. Both *state-based* and *language-based* approaches aim to constrain a Petri net by adding places deduced from the set of *regions*. By now, many variations have been proposed, e.g., approaches dealing with multiple tokens in a place, region definitions for Petri nets with inhibitor arcs, extensions to partial languages, regions for infinite languages, etc.

Initially, region theory focused on *synthesis*. We require the input and the behavior of the resulting Petri net to be equivalent. Recently, region-based research started to focus on *process mining* as well where the goal is *not* to create an equivalent model but to *infer* new knowledge from the input. Process mining examines observed behavior rather than assuming a complete description in terms of a transition system or prefix-closed language. For this reason, one needs to deal with new problems such as noise and incompleteness. Equivalence notions are replaced by trade-offs between fitness, simplicity, precision, and generalization. A model with good *fitness* allows for most of the behavior seen in the event log. A model that does not *generalize* is “overfitting”. Overfitting is the problem that a very specific model is generated whereas it is obvious that the log only holds example behavior. A model that allows for “too much behavior” lacks *precision*. Simplicity is related to Occam’s Razor which states that “one should not increase, beyond what is necessary, the number of entities required to explain anything”. Following this principle, we look for the *simplest* process model that can explain what was observed in the event log. Process discovery from event logs is very challenging because of these and many other trade-offs. Clearly, there are many theoretical process-mining challenges with a high practical relevance that need to be addressed urgently.

All these challenges and opportunities are the motivation for organizing the *Algorithms & Theories for the Analysis of Event Data* (ATAED) workshop. The workshop first took place in Brussels in 2015 as a succession of the *Applications of Region Theory* (ART) workshop series. From there on, the workshop moved to Toruń (2016), Zaragoza (2017), Bratislava (2018), and Aachen (2019). After the success of these workshops, it is only natural to bring together researchers working on region-based synthesis and process mining again.

The ATAED’2020 workshop took place as a virtual workshop on June 24, 2020 and was a satellite event of the 41st International Conference on Application and Theory of Petri Nets and Concurrency (Petri Nets 2020). We planned the workshop to be held in Paris, but due to the covid19 outbreak it became a virtual workshop.

Papers related to process mining, region theory and other synthesis techniques were presented as prerecorded videos at the ATAED’2020. These techniques have in common that “lower level” behavioral descriptions (event logs, partial languages, transition systems, etc.) are used to create “higher level” process

models (e.g., various classes of Petri nets, BPMN, or UML activity diagrams). In fact, all techniques that aim at learning or checking concurrent behavior from transition systems, runs, or event logs were welcomed. The workshop was supported by the IEEE Task Force on Process Mining (www.tf-pm.org/).

After a careful reviewing process, five papers were accepted for the workshop. We thank the reviewers for providing the authors with valuable and constructive feedback. We thank the authors and the presenters for their wonderful contributions.

Enjoy reading the proceedings!

Wil van der Aalst, Robin Bergenthum, and Josep Carmona
June 2020

Program committee of ATAED'2020

Wil van der Aalst, RWTH Aachen, Germany (co-chair)
Abel Armas Cervantes, QUT, Australia
Robin Bergenthum, FernUni Hagen, Germany (co-chair)
Luca Bernardinello, Università degli studi di Milano-Bicocca, Italy
Andrea Burattin, University of Innsbruck, Austria
Josep Carmona, UPC Barcelona, Spain (co-chair)
Claudio Di Ciccio, Vienna University of Economics and Business, Austria
Benoît Depaire, Hasselt University, Belgium
Jörg Desel, FernUni Hagen, Germany
Dirk Fahland, TU Eindhoven, The Netherlands
Chiara Di Francescomarino, FBK-IRST, Italy
Stefan Haar, LSV CNRS & ENS de Cachan, France
Gabriel Juhás, Slovak University of Technology, Slovak Republic
Anna Kalenkova, Higher School of Economics NRU, Russia
Jetty Kleijn, Leiden University, The Netherlands
Manuel Lama, University of Santiago de Compostela, Spain
Wen Lijie, Tsinghua University, China
Robert Lorenz, Uni Augsburg, Germany
Marta Pietkiewicz-Koutny, Newcastle University, GB
Adrian Puerto Aubel, INRIA - Rennes, France
Arik Senderovich, Technion, Israel
Jochen De Weerd, KU Leuven, Belgium
Moe Wynn, Queensland University of Technology, Australia
Alex Yakovlev, Newcastle University, GB

Table of Contents

Mitchel Brunings, Dirk Fahland, Boudewijn van Dongen <i>Defining meaningful Local Process Models</i>	6 - 19
Lisa Mannel, Robin Bergenthum, Wil van der Aalst <i>Removing Implicit Places Using Regions for Process Discovery</i>	20 - 32
Lisa Petrak, Robert Lorenz <i>Detecting Infrequent Behavior in Event Logs using Statistical Inference</i>	33 - 48
Pieter Kwantes, Jetty Kleijn <i>On Discovering Distributed Process Models the case of asynchronous communication</i>	49 - 65
Ronny Tredup, Evgeny Erofeev <i>On the Complexity of Synthesis of nop-Free Boolean Petri Nets</i>	66 - 84