

Workshop Proceedings

**Workshop on
Algorithms & Theories for the
Analysis of Event Data (ATAED'2015)**

Brussels, Belgium, June 22-23, 2015

Satellite event of the conferences

**15th International Conference on Application of
Concurrency to System Design (ACSD 2015)**

**36th International Conference on Application and Theory
of Petri Nets and Concurrency (PN 2015)**

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

These proceedings are published online by the editors as Volume 1371 at CEUR
Workshop Proceedings

<http://ceur-ws.org/Vol-1371>

Copyright © 2015 for the individual papers is held by the papers authors. Copy-
ing is permitted only for private and academic purposes. This volume is published
and copyrighted by its editors.

Preface

Regions have been defined about 20 years ago by Ehrenfeucht and Rozenberg as sets of nodes of a finite transition system that correspond to potential conditions that enable or disable transition occurrences in a corresponding elementary net system. Later, similar concepts were used to derive Petri nets from languages. Both *state-based* and *language-based* approaches aim to constrain a Petri net by adding places that are called *regions*. Over time, many variations have been proposed, e.g., approaches dealing with multiple-token in a place, extensions to partial orders, etc.

Initially, region theory focused on *synthesis* where the input behavior and resulting Petri net are supposed to be equivalent with respect to some equivalence criterion (e.g., bisimilar). Recently, region-based research started to focus also on *process mining* where the goal is *not* to create an equivalent model, but to *infer* new knowledge from the input. Process mining takes as input observed behavior rather than assuming a complete description in terms of a transition system or prefix-closed language. Classical region based techniques are unable to discover process models from event logs. 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* (i.e., is underfitting). 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 often 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.

The challenges and opportunities formed the main motivation for proposing the *Algorithms & Theories for the Analysis of Event Data* (ATAED’2015) workshop as a succession of the *Applications of Region Theory* (ART) workshop series. Our goal was (and is) to bring together researchers working on region-based synthesis and process mining. Looking at the proceedings, we succeeded in doing so!

The ATAED’2015 workshop took place in Brussels on June 22-23, 2015 and was a satellite event of both the 36th International Conference on Application and Theory of Petri Nets and Concurrency (Petri nets 2015) and the 15th International Conference on Application of Concurrency to System Design (ACSD 2015). Papers related to process mining, region theory and other synthesis techniques were presented at ATAED’2015. These techniques have in common that ‘lower level’ behavioral descriptions (event logs, partial orders, 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.win.tue.nl/ieeetfpm/).

After a careful reviewing process, eleven papers were accepted for the workshop. Overall, the quality of the submitted papers was good and most submissions matched very well the workshop goals. We thank the reviewers for providing the authors with valuable and constructive feedback. Moreover, we were honored that *Eike Best* (University of Oldenburg) was willing to give an invited talk on the “*Synthesis of Diamonds*”. We thank Eike, the authors, and the presenters for their wonderful contributions.

In the remainder, the accepted papers of the Algorithms & Theories for the Analysis of Event Data (ATAED’2015) workshop are briefly summarized.

- The paper “*On Binary Words Being Petri Net Solvable*” by Kamila Barylska, Eike Best, Evgeny Erofeev, Lukacs Mikulski, and Marcin Piatkowski studies the class of two-letter Petri net solvable words, i.e., Petri nets with two transitions and a reachability graph isomorphic to a trace-based transition system. Several intriguing results are presented for this class, e.g., the existence of side-place-free solutions given particular conditions.
- Andrey Mokhov and Josep Carmona use Conditional Partial Order Graphs (CPOGs) to create compact and easy-to-comprehend visualizations of event logs with data. In their paper “*Event Log Visualization with Conditional Partial Order Graphs: From Control Flow to Data*”, the authors provide a technique to automatically derive the control-flow part of the CPOG representation from an event log, and then incorporate the data contained in the log as conditions for the CPOG vertices and arcs.
- The paper “*Discovery of Personal Processes from Labeled Sensor Data: An Application of Process Mining to Personalized Health Care*” by Timo Szttyler, Johanna Völker, Josep Carmona, Oliver Meier, and Heiner Stuckenschmidt shows how process mining can be used for analyzing self-tracking data. Smart-phones and smart-watches can be used to produce detailed data about someone’s daily life. The authors describe the acquisition of such data in real-life and use existing process mining techniques for eliciting, analyzing and monitoring daily routines.
- “*ILP-Based Process Discovery Using Hybrid Regions*” by Sebastiaan van Zelst, Boudewijn Van Dongen, and Wil van der Aalst unifies the two existing types of language-based regions (single variable-based regions and dual variable-based regions) to provide a representation suitable for process mining. Integer Linear Programming (ILP)-based process discovery is further enhanced with a generalized ILP objective function. It is shown that any instantiation of the objective function leads to ILPs that favor minimal regions.
- Robin Bergenthum, Thomas Irgang, and Benjamin Meis present a folding algorithm to construct a business process model from a specification in their paper “*Folding Example Runs to a Workflow Net*”. Different to mainstream process mining techniques the input is not a sequential event log but a set of

example runs represented as labeled partial orders. By adopting ideas from the theory of regions, the authors aim at improving precision of the model while folding the runs into a model.

- The paper “*Mining Duplicate Tasks from Discovered Processes*” by Borja Vázquez-Barreiros, Manuel Mucientes, and Manuel Lama tackles the classical problem of label splitting in process mining. The authors propose an approach that uses the local information in the log to enhance an already mined model by performing a local search over the potential tasks to be duplicated. Experimental results show that, in a case study, the final model was improved in 35 out of 36 cases.
- The paper “*A Method For Assessing Parameter Impact on Control-Flow Discovery Algorithms*” by Joel Ribeiro and Josep Carmona presents a method to automatically assess the impact of parameters of control-flow discovery algorithms. The metaheuristic approach for process mining can be used to guide the user in selecting a technique, representational bias, and suitable parameter setting. The method has been evaluated over a set of logs while using the flexible heuristic miner.
- Antonia Azzini, Paolo Ceravolo, Ernesto Damiani, and Francesco Zavatartelli introduce the notion of extended behavior in their paper “*Knowledge Driven Behavioural Analysis in Process Intelligence*”. They present a methodology where first the descriptive knowledge is collected and the knowledge base queried, then (in the prescriptive and predictive knowledge phases) business rules and objectives are evaluated and unexpected business patterns and exceptions are uncovered.
- “*Compact Regions for Place/Transition Nets*” by Robin Bergenthum presents an approach using compact regions to synthesize a Petri net from a partial language. The language of a marked Petri net is its set of compact valid example runs. Compact regions are relevant as they may lead to faster synthesis algorithms computing smaller Petri nets. Initial results suggest that synthesis is indeed faster and that the compact solution space leads to nets having less places.
- In “*An Optimal Process Model for a Real Time Process*”, the authors Likewin Thomas, Manoj Kumar M.V., Annappa B., and Vishwanath K.P. provide a solution for recommending an optimal path of execution taking into account resource allocations. The proposed AlfyMiner compares variants of the same process encountered in different organizations. The authors include functionality to compare processes and to analyze resource behavior. This is then used to recommend next actions and suitable resources.
- The paper “*Capturing the Sudden Concept Drift in Process Mining*” by Manoj Kumar M.V., Likewin Thomas and Annappa B. focuses on sudden changes during process execution, i.e., second-order process dynamics. The paper proposes the extraction of a so-called “event class correlation feature” from logs for localizing the sudden concept drift in the control-flow perspective of the operational process. Experiments using synthetic event data show that (under ideal circumstances) sudden process changes can be detected.

The workshop proceedings provide a nice selection of ongoing research on the intersection of process mining and region-based synthesis. The papers illustrate the range of problems and solution approaches related to lifting ‘lower level’ dynamic behavior to ‘higher level’ process models. Given the rapid growth of event data, the area is expected to become even more relevant in years to come. We hope that ATAED’2015 serves as a starting point for a viable workshop series bringing together the two communities working on process mining and region-based synthesis.

Enjoy reading the proceedings!

Wil van der Aalst, Robin Bergenthum, and Josep Carmona
June 2015

Program committee of ATAED’2015

Wil van der Aalst, TU Eindhoven, The Netherlands (co-chair)
Rafael Accorsi, Universitaet Freiburg, Germany
Eric Badouel, INRIA Rennes, France
Robin Bergenthum, FernUni Hagen, Germany (co-chair)
Luca Bernardinello, Universit degli studi di Milano-Bicocca, Italy
Seppe vanden Broucke, KU Leuven, Belgium
Benoît Caillaud, INRIA Rennes, France
Toon Calders, Universit Libre de Bruxelles, Belgium
Josep Carmona, UPC Barcelona, Spain (co-chair)
Paolo Ceravolo, University of Milan, Italy
Benoît Depaire, Hasselt University, Belgium
Jörg Desel, FernUni Hagen, Germany
Boudewijn van Dongen, TU Eindhoven, The Netherlands
Luciano Garca-Bañuelos, University of Tartu, Estonia
Luís Gomes, Universidade Nova de Lisboa, Portugal
Gabriel Juhás, Slovak University of Technology, Slovak Republic
Anna Kalenkova, Higher School of Economics NRU, Russia
Jetty Kleijn, Leiden University, The Netherlands
Robert Lorenz, Uni Augsburg, Germany
Zbigniew Paszkiewicz, PricewaterhouseCoopers, Belgium
Marta Pietkiewicz-Koutny, Newcastle University, GB
Grzegorz Rozenberg, Leiden University, The Netherlands
Marcos Sepúlveda, Universidad Catolica de Chile, Chile
Jianmin Wang, Tsinghua University, China
Jochen De Weerd, KU Leuven, Belgium
Alex Yakovlev, Newcastle University, GB

Table of Contents

K. Barylska, E. Best, E. Erofeev, L. Mikulski, M. Piatkowski <i>On Binary Words Being Petri Net Solvable</i>	1 - 15
A. Mokhov, J. Carmona <i>Event Log Visualisation with Conditional Partial Order Graphs from Control Flow to Data</i>	16 - 30
T. Sztyley, J. Völker, J. Carmona, O. Meier, H. Stuckenschmidt <i>Discovery of Personal Processes from Labeled Sensor Data - An Application of Process Mining to Personalized Health Care</i>	31 - 46
S.J. van Zelst, B.F. van Dongen, W.M.P. van der Aalst <i>ILP-Based Process Discovery Using Hybrid Regions</i>	47 - 61
R. Bergenthum, T. Irgang, B. Meis <i>Folding Example Runs to a Workflow Net</i>	62 - 77
B. Vázquez-Barreiros, M. Mucientes, M. Lama <i>Mining Duplicate Tasks from Discovered Processes</i> (short paper)	78 - 82
J. Ribeiro, J. Carmona <i>A Method for Assessing Parameter Impact on Control-Flow Discovery Algorithms</i>	83 - 96
A. Azzini, P. Ceravolo, E. Damiani, F. Zavatarelli <i>Knowledge Driven Behavioural Analysis in Process Intelligence</i>	97 - 111
R. Bergenthum <i>Compact Regions for Place/Transition Nets</i> (short paper)	112 - 116
Likewin Thomas, Manoj Kumar M V, Annappa B, Vishwanath K P <i>An Optimal Process Model for a Real Time Process</i>	117 - 131
Manoj Kumar M V, Likewin Thomas, Annappa B <i>Capturing the Sudden Concept Drift in Process Mining</i>	132 - 143