

# Lecture Notes in Business Information Processing

203

Series Editors

Wil van der Aalst

*Eindhoven Technical University, Eindhoven, The Netherlands*

John Mylopoulos

*University of Trento, Povo, Italy*

Michael Rosemann

*Queensland University of Technology, Brisbane, QLD, Australia*

Michael J. Shaw

*University of Illinois, Urbana-Champaign, IL, USA*

Clemens Szyperski

*Microsoft Research, Redmond, WA, USA*

More information about this series at <http://www.springer.com/series/7911>

Paolo Ceravolo · Rafael Accorsi  
Philippe Cudre-Mauroux (Eds.)

# Data-Driven Process Discovery and Analysis

Third IFIP WG 2.6, 2.12 International  
Symposium, SIMPDA 2013  
Riva del Garda, Italy, August 30, 2013  
Revised Selected Papers

*Editors*

Paolo Ceravolo  
Università degli Studi di Milano  
Crema  
Italy

Philippe Cudre-Mauroux  
University of Fribourg  
Fribourg  
Switzerland

Rafael Accorsi  
Telematics  
University of Freiburg  
Freiburg  
Germany

ISSN 1865-1348                      ISSN 1865-1356 (electronic)  
Lecture Notes in Business Information Processing  
ISBN 978-3-662-46435-9              ISBN 978-3-662-46436-6 (eBook)  
DOI 10.1007/978-3-662-46436-6

Library of Congress Control Number: 2015932870

Springer Heidelberg New York Dordrecht London  
© IFIP International Federation for Information Processing 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media  
([www.springer.com](http://www.springer.com))

# Preface

The rapid growth of organizational and business process managed via information systems made available a big variety of data, which as a consequence created a high demand for making available data analysis techniques more effective and valuable. The third edition of the International Symposium on Data-driven Process Discovery and Analysis (SIMPDA 2013) was conceived to offer a forum where researchers from different communities and the industry can share their insights in this hot new field. The symposium features a number of advanced keynotes illustrating new approaches, presentations on recent research, a competitive PhD seminar, and selected research and industrial demonstrations. The goal is fostering exchanges among academic researchers, industry, and a wider audience interested on process discovery and analysis. The event is organized jointly by the IFIP WG 2.6 and W.G 2.12. In 2015 the symposium was inserted among the VLDB 2013 workshops, in the charming setting of Riva del Garda at the northwestern corner of Lake Garda, at the southern edge of the Italian Alps, near the Dolomites.

Submissions cover theoretical issues related to process representation, discovery, and analysis or provide practical and operational experiences in process discovery and analysis. To improve the quality of the contributions the symposium fostered the discussion during the presentation, letting authors the opportunity to improve their work extending the presented results. For this reason, authors of accepted papers and keynote speakers were invited to submit extended articles to this post-symposium volume of LNBIP. There were 18 submissions and 6 papers were accepted for publication.

In the first paper “The Effect of Noise on Mined Declarative Constraints,” Claudio Di Ciccio et al. present an experimental analysis on the robustness to noise of the different types of Declare constraints, establishing a constraint hierarchy based on formal relationships between the different types.

In the second paper by Gregor Grambow et al. “Towards Collecting Sustainability Data in Supply Chains with Flexible Data Collection Processes,” a dedicated information system for supporting data collection processes is developed. In combination with potentially long-running processes, these issues result in great process variability.

The third paper by Helen Balinsky et al., “Handling Environment for Publicly Posted Composite Documents,” focuses on Publicly Posted Composite Documents (PPCD) – a multipart document format retaining parts in their original formatting for usability, user convenience, and information availability, while providing document security and built-in access control for inter- and intra-organizational document workflows distributed over low-security channels. In particular, this paper illustrates a methodology to use OLE Automation, Add-Ons, and System Call Interception technique to preserve the overall document composition, while the handling of individual parts is delegated to external tools, native for individual document formats.

The fourth paper by Roberto Espinosa et al., “Enabling Non-expert Users to Apply Data Mining for Bridging the Big Data Divide,” offers to nonexpert miners a tool that just by uploading their data sets, returns them the more accurate mining pattern without dealing with algorithms or settings, thanks to the use of a data mining algorithm recommender. The authors also incorporate a previous task to help nonexpert users to specify data mining requirements and a later task in which users are guided in interpreting data mining results. Furthermore, they experimentally test the feasibility of our approach, in particular, the method to build recommenders in an educational context, where instructors of e-learning courses are nonexpert data miners who need to discover how their courses are used in order to make informed decisions to improve them.

The fifth paper by Antonia Azzini et al., “Combining Semantic Lifting and Ad Hoc Contextual Analysis in a Data Loss Scenario,” investigates the role of semantic lifting in discovering behavioral patterns of the executed process. Conclusions spotlight the impact that semantics lifting has on support and confidence of the inferred probabilities of observing these behavioral patterns.

The sixth paper by Wil M.P. van der Aalst et al., “Comparative Process Mining in Education: An Approach Based on Process Cubes,” starts by observing that existing process mining techniques focus on the analysis of a single process rather than the comparison of different processes. This paper proposes comparative process mining using process cubes. An event has attributes referring to the dimensions of the process cube. Through slicing, dicing, rolling-up, and drilling-down we can view event data from different angles and produce process mining results that can be compared.

We gratefully acknowledge the strong research community that gathered around the research problems related to process data analysis and the high quality of their research work, which is hopefully reflected in the papers of this issue. We also would like to express our deep appreciation for the referees’ hard work and dedication. Above all, thanks are due to the authors for submitting the best results of their work to the Symposium on Data-driven Process Discovery and Analysis.

We are very grateful to the Università degli Studi di Milano and to IFIP for their financial support, and to the University of Freiburg and the University of Fribourg.

November 2014

Paolo Ceravolo  
Rafael Accorsi  
Philippe Cudre-Mauroux

# Organization

## Conference Co-chairs

Paolo Ceravolo	Università degli Studi di Milano, Italy
Rafael Accorsi	University of Freiburg, Germany
Philippe Cudre-Mauroux	University of Fribourg, Switzerland

## Advisory Board

Karl Aberer	EPFL, Switzerland
Ernesto Damiani	Università degli Studi di Milano, Italy
Tharam Dillon	La Trobe University, Australia
Dragan Gasevic	Athabasca University, Canada
Marcello Leida	Etisalat BT Innovation Centre, UAE
Erich Neuhold	University of Vienna, Austria
Maurice van Keulen	University of Twente, The Netherlands

## PhD Award Committee

Gregorio Piccoli	Zucchetti spa, Italy
Paolo Ceravolo	Università degli Studi di Milano, Italy
Marcello Leida	Etisalat BT Innovation Centre, UAE

## Web Chair and Publicity Chair

Fulvio Frati	Università degli Studi di Milano, Italy
--------------	---

## Program Committee

Peter Spyns	Free University of Brussels, Belgium
Irene Vanderfeesten	Eindhoven University of Technology, The Netherlands
Daniele Bonetta	Università della Svizzera Italiana, Switzerland
Etienne Rivière	Université de Neuchâtel, Switzerland
Sylvain Hallé	Université du Québec à Chicoutimi, Canada
Ioana Georgiana	Free University of Brussels, Belgium
Schahram Dustdar	Vienna University of Technology, Austria
Hong-Linh Truong	Vienna University of Technology, Austria
Ebrahim Bagheri	Ryerson University, Canada
Mustafa Jarrar	Birzeit University, Palestinian Territory
Hamid Motahari	HP Labs, USA

Valentina Emilia Balas	University of Arad, Romania
George Spanoudakis	City University London, UK
Gregorio Martinez Perez	University of Murcia, Spain
Wei-Chiang Hong	Oriental Institute of Technology, Taiwan
Mohamed Mosbah	University of Bordeaux, France
Jerzy Korczak	Wroclaw University of Economics, Poland
Jan Mendling	Wirtschaftsuniversität Wien, Austria
Maurice van Keulen	University of Twente, The Netherlands
Bernhard Bauer	University of Augsburg, Germany
Christos Kloukinas	City University London, UK
Gabriele Ruffatti	Engineering Group, Italy
Alessandra Toninelli	Engineering Group, Italy
Eduardo Fernandez-Medina	University of Castilla-La Mancha, Spain
Chi Hung	Tsinghua University, China
Nora Cuppens	Telecom Bretagne, France
Debasis Giri	Haldia Institute of Technology, India
Wil van der Aalst	Technische Universiteit Eindhoven, The Netherlands
Antonio Mana Gomez	University of Malaga, Spain
Davide Storelli	Università del Salento, Italy
Jose M. Alcaraz Calero	Hewlett-Packard Labs, UK
Richard Chbeir	University of Bourgogne, France
Mohamed Achemlal	Orange Labs, France
Helen Balinsky	Hewlett-Packard Labs, UK
Karima Boudaoud	Ecole Polytechnique de Nice Sophia Antipolis, France
Meiko Jensen	University, Germany
Renato Iannella	Semantic Identity, Australia
Farookh Hussain	University of Technology, Sydney, Australia
Marcello Leida	Etisalat BT Innovation Centre, UAE
Frédéric Cuppens	Télécom Bretagne, France
Gerd Groner	University of Koblenz and Landau, Germany
Abder Koukam	University of Technology of Belfort-Montbéliard, France
Wei-Chiang Hong	Oriental Institute of Technology, Taiwan



# Contents

The Effect of Noise on Mined Declarative Constraints. . . . .	1
<i>Claudio Di Ciccio, Massimo Mecella, and Jan Mendling</i>	
Towards Collecting Sustainability Data in Supply Chains with Flexible Data Collection Processes . . . . .	25
<i>Gregor Grambow, Nicolas Mundbrod, Jens Kolb, and Manfred Reichert</i>	
Handling Environment for Publicly Posted Composite Documents. . . . .	48
<i>Helen Balinsky and David Subirós Pérez</i>	
Enabling Non-expert Users to Apply Data Mining for Bridging the Big Data Divide . . . . .	65
<i>Roberto Espinosa, Diego García-Saiz, Marta Zorrilla, Jose Jacobo Zubcoff, and Jose-Norberto Mazón</i>	
Combining Semantic Lifting and Ad-hoc Contextual Analysis in a Data Loss Scenario . . . . .	87
<i>Antonia Azzini, Ernesto Damiani, and Francesco Zavatarelli</i>	
Comparative Process Mining in Education: An Approach Based on Process Cubes . . . . .	110
<i>Wil M.P. van der Aalst, Shengnan Guo, and Pierre Gorissen</i>	
<b>Author Index . . . . .</b>	<b>135</b>