



# Computational modelling and data-driven techniques for systems analysis

Stan Matwin<sup>1,2</sup>  · Luca Tesei<sup>3</sup> · Roberto Trasarti<sup>4</sup>

Received: 7 March 2019 / Accepted: 8 March 2019 /

Published online: 11 April 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## 1 Introduction

This JIIS Special Issue aimed at bringing together contributions from academia, industry and research institutions interested in the combined application of computational modelling methods with data-driven techniques from the areas of knowledge management, data mining and machine learning. Modelling methodologies of interest included automata, agents, Petri nets, process algebras and rewriting systems. Application domains included social systems, ecology, biology, medicine, smart cities, governance, education, software engineering, and any other field that deals with complex systems and large amounts of data.

Contributions were open to the participants of the DataMod 2016 symposium and its previous editions (<http://pages.di.unipi.it/datamod/edition-2016/>) and to the general audience. The 5th International Symposium: From Data to Models and Back (DataMod 2016) former MoKMaSD symposium (the International Symposium on Modelling and Knowledge Management: Systems and Domains) was organised as a satellite event of STAF 2016 (associated to SEFM). MoKMaSD was a symposium dealing with approaches that integrate modelling and knowledge management/discovery methods for the study of complex systems in almost any application domain. The symposium name was changed to improve the fit with the current interests of the community that has been constituted over the years around

---

✉ Luca Tesei  
luca.tesei@unicam.it

Stan Matwin  
stan@cs.dal.ca

Roberto Trasarti  
roberto.trasarti@isti.cnr.it

<sup>1</sup> Faculty of Computer Science, Dalhousie University, 6050 University Ave., P.O. Box 15000, Halifax, Nova Scotia, B3H 4R2, Canada

<sup>2</sup> Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

<sup>3</sup> School of Science and Technology, University of Camerino, via Madonna delle Carceri 9, 62032 Camerino, Italy

<sup>4</sup> ISTI-CNR, National Research Council of Italy, via G. Moruzzi 1, 56124 Pisa, Italy

it. All of the four previous editions of the symposium were organised as satellite events of SEFM (in 2012, 2013, 2014 and 2015). The symposium's success has seen consistent growth over the years.

The guest editors Luca Tesei and Roberto Trasarti served as Program Chairs for DataMod 2016. Four papers in this special issue are extended versions of papers presented at DataMod 2016 and one is an extended version of a paper presented at MoKMaSD 2015. They were originally peer-reviewed by the MoKMaSD 2015 and DataMod 2016 program committees and the extended versions undertook another process of peer-reviewing for the special issue. Moreover, there are two papers that have been newly submitted to the special issue. They undertook the same process of peer-reviewing by selected experts from the relative areas.

## 2 Contributions in this special issue

The papers in this Special Issue focus on computational and data-driven modelling techniques and tools. Also, new data analysis techniques have been proposed.

In “Improving process algebra model structure and parameters in infectious disease epidemiology through data mining” by Hamami, Atmani, Cameron, Pollock and Shankland, particular data mining techniques are used to improve the modelling process of a health-related phenomenon. They demonstrate the approach with an epidemiological Bio-PEPA model for the mumps virus. “Model Mining - Integrating Data Analytics, Modelling and Verification” by Cerone is an extension of a DataMod 2016 paper that introduces the notion of model mining, an improvement of available process mining techniques in the business process management area to extract business processes from event logs. The model that is derived is a set of formal rules for generating the system behaviour and is more abstract and concise of models that result from the existing process mining techniques. The technique is applied in two different case studies, one in the field of ecology and the other in the field of collaborative learning. “Objective/MC: A High-Level Model Checking Language. Formalisation of the Imperative Core and Translation into PRISM” by Milazzo and Pardini, an extension of a paper presented at DataMod 2016, proposes a high-level language for modelling finite-state systems that is designed for being usable by non-experts, enabling them to use model checking verification effectively. The paper also presents a compiler from the language to the model checking tool PRISM.

“Measuring Network Reliability and Repairability against Cascading Failures”, by Thapa, Espejo-Urbe and Pournaras, introduces a probabilistic framework for measuring network reliability and repairability against cascading failures in critical infrastructures like power grids, water/gas networks, economic markets, traffic systems and so on. With the experimental evaluation it is shown that the generic measurements of the framework improve the understanding of reliability and repairability in systems of different nature. “Design of a Software Architecture Supporting Business-to-Government Information Sharing to Improve Public Safety and Security: Combining Business Rules, Events and Blockchain Technology” by van Engelenburg, Janssen and Klievink is an extension of a paper presented at MoKMaSD 2015. The designed architecture is a blockchain that stores events and rules for information sharing. The objective is to meet both the need of governments to collect information about transported goods for security reasons and the need of companies to keep information confidential.

“One-pass MapReduce-based Clustering Method for Mixed Large Scale Data” by HajKacem, N’Cir and Essoussi proposes a new method that largely reduces the input/output operations required by the MapReduce implementation of k-prototypes. It also

accelerates the clustering process using a pruning strategy that eliminates redundant computations. Experiments show that the method is scalable and improve the efficiency of existing methods. This paper is an extension of a paper presented at DataMod 2016. “Persistent Entropy for Separating Topological Features from Noise in Vietoris-Rips Complexes” by Gonzalez-Diaz, Atienza and Rucco, originally presented at DataMod 2016, presents new important properties of the notion of persistent entropy, a measure introduced in the emerging field of topological data analysis. The introduced properties are also used to introduce a simple method to separate topological noise from real features in Vietoris-Rips filtrations.

The papers in this Special Issue, to different extent, introduce or use methods and techniques that are data-driven, combining them with different notions of computational models. We hope that this fruitful integration will be further developed in future works improving cross fertilisation among the research communities of modelling and data analysis.

**Acknowledgements** DataMod 2016 and this work were supported by the Future and Emerging Technologies (FET) program within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FET-Proactive grant agreement TOPDRIM (<http://www.topdrim.eu>), number FP7-ICT-318121. This work was also supported by the European Community’s H2020 Program under the scheme “INFRAIA-1-2014-2015: Research Infrastructures”, grant agreement n. 654024 “SoBigData: Social Mining & Big Data Ecosystem” (<http://www.sobigdata.eu>).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.