EDITORIAL

Editorial special issue on data analytics for engineering, science and society

Tania Cerquitelli[1] ⓘ · Silvia Chiusano[2] ⓘ · Genoveva Vargas-Solar[3] ⓘ

Today a wide range of technological and pervasive applications have made available a huge amount of heterogeneous complex data collections. Digging deep these data collections can unearth a rich spectrum of knowledge valuable to add intelligence in target domains. However, data analytics performed on such collections is still a daunting task, because data are generally too dirty, complex, and heterogeneous to be processed through state-of-the-art data analysis techniques. Consequently, various data science challenges arise, tailored to each application domain, related with the creation, storage, search, sharing, modelling, analysis, and visualisation of data and associated inner knowledge.

Algorithms are powerful tools to manage and analyse data aimed to fulfill data science expectations. Even if people barely notice them, they are key and enabling components of the services used in their daily lives. Unfortunately most algorithms, even if they are precise and effective, remain obscure and human unreadable. Making algorithms' choices and outcomes understandable and transparent would allow people to have insight about the background of algorithms' choices, and this aspect could contribute to improve their usability in various real-life settings. Transparent solutions are thus needed to produce more credible and reliable data analytics services, playing a key role in proactive user engagement by making widely accessible the results of the data analytical processes.

This special issue introduces original contributions aiming to make data analytics models and technique accessible to people and useful to address engineering,

✉ Tania Cerquitelli
    tania.cerquitelli@polito.it

Silvia Chiusano
silvia.chiusano@polito.it

Genoveva Vargas-Solar
genoveva.vargas@imag.fr

[1] Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

[2] Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, Turin, Italy

[3] LIG-LAFMIA, French Council of Scientific Research (CNRS), Grenoble, France

scientific and societal challenges. Each contribution included in the special issue describes interesting research activities developed in the field of applied data science innovation. These contributions have a wide variety of perspectives including methodological aspects, theoretical and technological views. The special issue covers a wide range of interesting topics categorised into: (1) *solutions addressing issues related to exploit/deploy data science technologies* such as load scheduling in distributed environments (Bhatia et al.), quality of service (Haytamy et al.), data retrieval strategy (Andrejev et al.), reduction of data transmissions in sensor networks (Kniess and Oliviera), privacy in the data analytics workflow (Belhajjame et al.); (2) *innovative analytics algorithms towards interpretable models* proposed to add more transparency in some application scenarios such as Healthcare (Fawagreh and Gaber), electric distribution networks (Renga et al.), quantitative stock trading (Cagliero et al.); (3) *characterizing user behavior and the surrounding environments* as discussed by Chen and Golab in the context of micro-journal mining to understand environment stimuli. Boratto et al. present a strategy to model user behavior in journey planning, while Castells-Graells et al. discuss how to address cyclings' urban safety. Finally, Jabbour et al. present a methodology to easily detect overlapping community in networks.

More in detail, the special issue starts with three research contributions discussing different topics related to effectively provide services by exploiting data science technologies.

Bhatia et al. in *Quantumized Approach of Load Scheduling in Fog Computing Environment for IoT Applications* discuss a quantumized approach for scheduling heterogeneous tasks in the context of fog computing environments to support the decision-making process in real-time. A comparative analysis has been also presented to show the effectiveness of the proposed approach with respect to state-of-the-art methodologies.

Haytamy et al. in *A Deep Learning based Framework for Optimizing Cloud Consumer QoS-based Service Composition* introduce a combined approach, tested on a real dataset, to optimize cloud consumer QoS-based Service composition exploiting the deep learning long short term memory (LSTM) network and the particle swarm optimization (PSO) algorithm.

Andrejev et al. in *Strategies for Array Data Retrieval from a Relational Back-End Based on Access Patterns* propose a framework to efficiently process array data retrieval queries to support different data processing tasks including adaptive pattern discovery and pre-fetching of chunks from external storage. Different strategies of array data retrieval are evaluated by considering data partitioning options and access patterns.

Kniess and Oliviera in *Data Reduction in Sensor Networks Based on Dispersion Analysis* present a strategy to reduce the number of data transmissions in sensor nodes. In order to maintain low processing time, the proposed method is based on data dispersion analysis to avoid transmitting measurements with low dispersion values. The experimental evaluation demonstrates the efficiency in reducing data transmissions in wireless sensor networks while keeping low CPU utilization rate but good data accuracy.

Belhajjame et al. in *On Privacy-Aware eScience Workflows* first discuss research issues of privacy-aware Data analysis Workflow (DWf) and then propose a 2-step

strategy that automatically identifies the sensitivity and anonymity-degree of the DWf parameters. This study also proposes a strategy to anonymize data produced by DWf executions allowing third-party scientists to verify the claims made by the authors in the scholarly articles.

The special issue continues with the discussion of cutting-edge data science solutions to enhance the interpretable aspect of predictive models.

Fawagreh and Gaber in *Resource-efficient Fast Prediction in Healthcare Data Analytics: A Pruned Random Forest Regression Approach* investigate a method to build a prediction model with constrained-resources, as an extension of the Random Forest algorithm with similar performance but smaller in size. The proposed approach relies on the cluster analysis and has been tested in the context of healthcare data analysis by discussing results on datasets related to patients affected by three diseases.

Renga et al. in *Data-driven exploratory models of an electric distribution network for fault prediction and diagnosis* present a data science pipeline to automatically discover interesting and human-readable insights in the context of a real-world medium-voltage distribution. Experimental results, obtained on a real-dataset, demonstrate the effectiveness of the proposed approach in discovering interesting correlations from SCADA (Supervisory Control And Data Acquisition) events that could be used for prognostic-diagnostic purposes.

Cagliero et al. in *Training ensembles of faceted classification models for quantitative stock trading* propose a new ensemble method to forecast future stock prices, which combines predictions based on separate sets of features characterizing the stock data. Obtained outputs are then properly combined to produce a multi-faceted recommendation.

Three research papers conclude the issue with an interesting discussion on different aspects related to user behaviour and environment modeling.

Chen and Golab in *Micro-Journal Mining to Understand Mood Triggers* study environment stimuli that contribute to different moods by analyzing a large collection of micro-journalling. The proposed approach relies first on a multimodal logistic regression to classify micro-journal with a mood label and then on a subsequent interpretation of the predictive model through trigrams.

Boratto et al. in *Characterizing user behavior in journey planning* address the analysis of the user behaviors in journey planning from the viewpoint of the perceived value of travel time. The study is performed in an international context, considering users from several countries and multiple factors jointly affecting user behavior (i.e., search, sorting, and selection).

Castells-Graells et al. in *On Cycling Risk and Discomfort: Urban Safety Mapping and Bike Route Recommendations* present a personalized route recommendation system towards cyclings' urban safety. The proposed solution relies on the analysis of data collected from different sources to map cycling risk in the area of Zurich and an estimation model that balances spatial cycling risk and route discomfort.

Jabbour et al. in *SAT-based Models for Overlapping Community Detection in Networks* propose a novel strategy to automatically discover k-centered linked communities with boundaries diameter. The proposed approach relies on a refined strategy to enhance the partial maximum satisfiability optimization problem. The ability of the

approach in detecting relevant communities in networks has been proven on target real-world networks.

The Guest Editors would like to express their thanks to all those who contributed to the success of the special issue. We thank all the authors attracted by the special issue topics who submitted their research contributions. We received 50 submitted papers and only 12 of them are accepted to be included in the issue. The accepted papers are written by authors developing their research activities in a wide geographical area including Brazil, Canada, France, Egypt, Italy, India, Spain, Sweden, Switzerland, and United Kingdom.

Special thanks go to the reviewers who offered their expertise and valuable time in evaluating the submitted papers, providing comprehensive, critical and constructive comments to the authors, and ensuring high-quality research standards of the special issue. Moreover, our deepest appreciation to the Computing Editor-In-Chief Prof. Schahram Dustdar, for his professional, precious and timely support and to the Springer staff, in particular to Linda Xavier and Christine Kamper for their hard and timely working.

We hope readers will find the special issue interesting and that the content will contribute to the development of successful research!

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.