PROCEEDINGS OF THE 2019 INTERNATIONAL CONFERENCE ON DATA SCIENCE

ICDATA'19

Editors

Robert Stahlbock

Gary M. Weiss, Mahmoud Abou-Nasr

Associate Editor

Hamid R. Arabnia

Copyright and Reprint Permission

American Council on Science and Education (ACSE)

Foreword

It gives us great pleasure to introduce this collection of papers to be presented at the 15th International Conference on Data Science 2019, ICDATA'19 (https://icdata.org), July 29 – August 1, 2019, at Luxor Hotel, Las Vegas, USA. The preliminary edition of this book (available in July 2019 for distribution on site at the conference) includes only a subset of the accepted research articles. The final edition (available in August 2019) will include all accepted research articles. This is due to deadline extension requests received from many authors who wished to continue enhancing the write-up of their papers (by incorporating the referees' suggestions). The final edition of the proceedings will be made available at https://americancse.org/events/csce2019/proceedings.

Data mining or machine learning is critically important if we want to effectively learn from the tremendous amounts of data that are routinely being generated in science, engineering, medicine, business, and other areas. The aim is gaining insight into processes, transactions, extract knowledge, make better decisions, and deliver value to users or organizations. This is even more important and challenging in an era in which scientists and practitioners are faced with numerous challenges caused by exponential expansion of digital data, its diversity and complexity. The scale and growth of data considerably outpace technological capacities of organizations to process and manage it. During the last decade, we all observe new, more glorious and promising concepts or labels emerging and slowly but steadily displacing 'data mining' from the agenda of CTO's. It was and still is the time of data science, big data, advanced-/business-/customer-/data-/predictive-/prescriptive-/…/risk-analytics, to name only a few terms that dominate websites, trade journals, and the general press – although there is even a rebirth of terms such as artificial intelligence (AI) and (machine) learning (e.g., deep learning) in academia, companies, and even on the agenda of political decision makers.

All the concepts of data science aim at leveraging data for a better understanding of complex real-world phenomena. They all pursue this objective using some formal, often algorithmic, procedures, at least to some extent. This is what data miners have been doing for decades. The very idea of all those similar or identical concepts with different labels, the idea to think of massive, omnipresent amounts of data as strategic assets, and the aim to capitalize on these assets by means of analytic procedures is, indeed, more relevant and topical than ever before. Although there are very helpful advances in hardware and software, there are still many challenges to be tackled in order to leverage the promises of data analytics. Obviously, technological change is never ending and appears to be accelerating. Right now the world seems especially focused on machine learning and data mining (not contradictory but similar or even equivalent to data science), as these disciplines are making an ever increasing impact on our society. Large multinational corporations are expanding their efforts in these areas and students are flocking to computer science and related disciplines in order to learn about these disciplines and take advantage of the many lucrative job opportunities. Many industries, even conservative ones like, e.g., the port industry, are working towards "Version 4.0" (e.g., "Port 4.0"), with digitization, digitalization and even digital transformation of traditional processes resulting in improved workflows, new concepts and new business plans. Their goal usually includes data analytics, automation, autonomization, robotics, AI etc. The industry is interested in feasibility studies and results of scientific research. Data science is popular like never before. Data scientists are rare on the job market and, therefore, very well payed.

The growth in all these areas has been dramatic enough to require changes in nomenclature. Most of these 'hot' technologies and methods are increasingly considered part of the broad field of data science, and there are benefits to viewing this field as a unified whole, rather than a collection of disparate sub-disciplines. ICDATA, the former data mining conference DMIN merged with the big data conference ABDA, is much broader than just data mining and big data. It includes all of the following main topics: all aspects of data mining and machine learning (tasks, algorithms, tools, applications, etc.), all aspects of big data (algorithms, tools, infrastructure, and applications), data privacy issues, and data management. The conference is designed to be of equal interest to researchers and practitioners, academics and members of industry, computer scientists, physical and social scientists, and business analysts.

Data science attracts innovative and influential contributions to both research and practice, across a wide range of academic disciplines and application domains. Our conference seeks to acknowledge and facilitate excellence in research and applications in the area of data science. Our conference is held annually under the umbrella of the World Congress in Computer Science, Computer Engineering, and Applied Computing (CSCE). An important mission of CSCE includes "*Providing a unique platform for a diverse community of*

constituents composed of scholars, researchers, developers, educators, and practitioners. The Congress makes concerted effort to reach out to participants affiliated with diverse entities (such as: universities, institutions, corporations, government agencies, and research centers/labs) from all over the world. The congress also attempts to connect participants from institutions that have teaching as their main mission with those who are affiliated with institutions that have research as their main mission. The congress uses a quota system to achieve its institution and geography diversity objectives." By any definition of diversity, this congress is among the most diverse scientific meeting in USA. We are proud to report that this federated congress has authors and participants from 57 different nations representing variety of personal and scientific experiences that arise from differences in culture and values. As can be seen (see below), the program committee of this conference as well as the program committee of all other tracks of the federated congress are as diverse as its authors and participants..

CSCE'19 assembles a spectrum of 20 affiliated research conferences, workshops, and symposiums into a coordinated research meeting. Each conference has its own program committee as well as referees and own indexed proceedings. Attendees have full access to all 20 conferences' sessions, tracks, and tutorials. ICDATA seeks to reflect the multi- and interdisciplinary nature of data science and to facilitate the exchange and development of novel ideas, open communication and networking amongst researchers and practitioners in different research domains. As in previous years, we hope that the 2019 International Conference on Data Science will provide a forum for you to present your research in a professional and open-minded environment, exchange ideas, and network and interact across research areas. ICDATA'19 provides an international and multicultural experience with contributions from 19 different countries. We consider the resulting diversity in attendees and the mixture of established and starting researchers as a particular advantage of an engaging conference format.

ICDATA'19 attracted submissions of theoretical research papers as well as industrial reports, application case studies, and in a second phase, late breaking papers, position papers, and abstract/poster papers. The program committee would like to thank all those who submitted papers for consideration. We strived to establish a review process of high quality. To ensure a fair, objective and transparent review process all review criteria are published on the website. Papers were evaluated regarding their relevance to ICDATA, originality, significance, information content, clarity, and soundness on an international level. Each aspect was objectively evaluated, with alternative aspects finding consideration for application papers. Each paper was refereed by at least two researchers in the topical area, taking the reviewers' expertise and confidence into consideration, with most of the papers receiving three reviews. The review process was competitive. The overall acceptance rate for submissions was 49%.

We are very grateful to the colleagues who helped in organizing the conference. In particular, we would like to thank the members of the program committee of ICDATA'19 and the members of the congress steering committee. The continuing support of the ICDATA program committee has been essential to further improve the quality of accepted submissions and the resulting success of the conference. The ICDATA'19 program committee members are (in alphabetical order): Mahmoud Abou-Nasr, Kai Brüssau, Paulo Cortez, Kevin Daimi, Antonio Dourado, Diego Galar, Peter Geczy, Tzung-Pei Hong, Wei-Chiang Hong, Ulf Johansson, Terje Solsvik Kristensen, Hossein Peyvandi, Zhang Sen, Robert Stahlbock, Vivian Sultan, Chamont Wang, Simon Wang, Gary M. Weiss, Zijiang Yang, and Yu Zhang. They all did a great job in evaluating a lot of submissions in short time. In particular, we want to thank Mahmoud Abou-Nasr for organizing the special session on "Real-World Data Mining & Data Science Applications, Challenges, and Perspectives" for more than a decade.

We would also like to thank our publicity co-chairs Ashu M. G. Solo (Fellow of British Computer Society, Principal/R&D Engineer, Maverick Technologies America Inc.) for circulating information on the conference, as well as www.KDnuggets.com, a platform for analytics, data mining and data science resources, for listing ICDATA'19.

Considering the increasing efforts of all towards the quality of the review process, the conference sessions and the social program of ICDATA'19, we are confident that you will find the conference stimulating and rewarding. It is a particular pleasure to provide data mining oriented invited talks and tutorials presented by the following esteemed members of the data mining community (in alphabetical order): Yuri Demchenko (University of Amsterdam, The Netherlands), Diego Galar (Luleå University of Technology, Sweden), Peter Geczy (AIST, Japan), Ulf Johansson (Jönköping University, Sweden) and Andrew H. Johnston (Mandiant, USA).

As Sponsors-at-large, partners, and/or organizers each of the followings (separated by semicolons) provided help for at least one track of the Congress: Computer Science Research, Education, and Applications Press (CSREA); US Chapter of World Academy of Science; American Council on Science & Education & Federated Research Council (http://www.americancse.org/). In addition, a number of university faculty members and their staff (names appear on the cover of the set of proceedings), several publishers of computer science and computer engineering books and journals, chapters and/or task forces of computer science associations/organizations from 3 regions, and developers of high-performance machines and systems provided significant help in organizing the conference as well as providing some resources. We are grateful to them all. We are also grateful for support by the Institute of Information Systems at Hamburg University, Germany.

We express our gratitude to keynote, invited, and individual conference/tracks and tutorial speakers - the list of speakers appears on the conference web site. We would also like to thank: Kaveh Arbtan and UCMSS (Universal Conference Management Systems & Support, California, USA) for managing all aspects of the conference; Dr. Tim Field of APC for coordinating and managing the printing of the proceedings; and the staff of Luxor Hotel (Convention department) at Las Vegas for the professional service they provided.

Last but not least, we wish to express again our sincere gratitude and utmost respect towards our colleague and friend Prof. Hamid R. Arabnia (Professor, Department of Computer Science, University of Georgia, USA; Editor-in-Chief, Journal of Supercomputing/Springer), General Chair and Coordinator of the federated congress, and also Associate Editor of ICDATA'19 for his excellent and tireless support, organization and coordination of all affiliated events. His exemplary and professional effort in 2019 and all the many years before in the steering committee of the congress makes these events possible. We are grateful to continue our data science conference as ICDATA'19 under the umbrella of the CSCE congress.

Thank you all for your contribution to ICDATA'19! We hope that you will experience a stimulating conference with many opportunities for future contacts, research and applications.

We present the proceedings of ICDATA'19.

Robert Stahlbock
ICDATA'19 General Conference Chair

Steering Committee ICDATA'19

https://icdata.org