PROCEEDINGS OF
THE 2016 INTERNATIONAL CONFERENCE ON
DATA MINING

DMIN 2016

Editors

Robert Stahlbock
Gary M. Weiss

Associate Editors

Mahmoud Abou-Nasr
Hamid R. Arabnia, Diego Galar
Peter Geczy

WORLDCOMP'16
July 25-28, 2016
Las Vegas Nevada, USA
www.worldcomp.org

This volume contains papers presented at The 2016 International Conference on Data Mining (DMIN'16). Their inclusion in this publication does not necessarily constitute endorsements by editors or by the publisher.

Copyright and Reprint Permission

ISBN: 1-60132-431-6
Printed in the United States of America

CSREA Press
U. S. A.

Foreword

It gives us great pleasure to introduce this collection of papers to be presented at the 12th International Conference on Data Mining 2016, DMIN’16 (www.dmin-2016.com), July 25-28, 2016, at Monte Carlo Resort, Las Vegas, Nevada, USA.

Data mining is a relatively young discipline that 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 in order to gain insight into processes, transactions, make better decisions, and deliver value to users or organizations.

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. This trend is expected to continue over the following years, bringing yet unforeseen challenges. However, during the last years, 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 is the time of data science, big data, advanced-/business-/customer-/data-/…/risk-analytics, to name only a few terms that dominate websites, trade journals, and the general press. But they all aim at leveraging data for a better understanding of, and insight into, 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. So maybe the label ‘data mining’ has lost much of its momentum and made room for more recent ‘competitors’, but the very idea of it, 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. Advances in hardware and software are helpful, but there are still many challenges to be tackled in order to leverage the promises of data analytics.

An important mission of the World Congress in Computer Science, Computer Engineering, and Applied Computing (a federated congress to which this conference is affiliated with) 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 74 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.

Data mining attracts innovative and influential contributions to both research and practice, across a wide range of academic disciplines and application domains. DMIN conferences seek to acknowledge and facilitate excellence in research and applications in the area of data mining. DMIN conferences are held annually within WORLDCOMP. WORLDCOMP'16 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. DMIN seeks to reflect the multi- and interdisciplinary nature of data mining 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 2016 International Conference on Data Mining will provide a forum for you to present your research in a professional environment, exchange ideas, and network and interact across research areas. DMIN actively supports students and beginning researchers from lesser developed countries by funding registration and accommodation, in order to allow for a truly international networking and understanding. DMIN’16 provides an international and multicultural experience with contributions from 22 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.

DMIN’16 attracted a high number of 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 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 were published on the website. Papers were evaluated regarding their relevance to DMIN, 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 paper acceptance rate for papers was 51%.

We are very grateful to the many colleagues who helped in organizing the conference. In particular, we would like to thank the members of the program committee of DMIN'16 and the members of the congress steering committee. The continuing support of the DMIN program committee has been essential to further improve the quality of accepted submissions and the resulting success of the conference. The DMIN'16 program committee members are (in alphabetical order): Mahmoud Abou-Nasr, Jérôme Azé, Alina Campan, Paulo Cortez, Kevin Daimi, Christian Dawson, Qin Ding, António Dourado, Philippe Fournier-Viger, Diego Galar, Peter Geczy, Zahid Halim, Tzung-Pei Hong, Wei-Chiang Hong, Ulf Johansson, Madjid Khalilian, Terje Kristensen, Philippe Lenca, Jerry Chun-Wei Lin, Wen-Yang Lin, Tanja Magoc, Rabie A. Ramadan, Gerald Schaefer, Sabrina Senatore, Victor Sheng, Yong Shi, Vijendra Singh, Robert Stahlbock, Chamont Wang, Simon Wang, Gary M. Weiss, Zijiang Yang, Yu Zhang, Songfeng Zhen, and Shang-Ming Zhou.

Many individuals listed above will be requested after the conference to provide their expertise and services for selecting papers for publication (extended versions) in journal special issues as well as for publication in a set of research books (to be prepared for publishers including: Springer, Elsevier, and others).

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 DMIN’16, and IEEE Intelligent Systems IS'16 as a listing partner (https://www.ieee-is.org/).

Considering the increasing efforts of all towards the quality of the review process, the conference sessions and the social program of DMIN'16, 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: Ulf Johansson (Jönköping University, Sweden), Diego Galar (Luleå University of Technology, Sweden), Peter Geczy (AIST, Japan), and Gary M. Weiss (Fordham University, NY, 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 (http://www.worldcomp.org/) ; American Council on Science & Education & Federated Research

Council (http://www.americancse.org/); HoIP, Health Without Boundaries, Healthcare over Internet Protocol, UK (http://www.hoip.eu); HoIP Telecom, UK (http://www.hoip-telecom.co.uk); and WABT, Human Health Medicine, UNESCO NGOs, Paris, France (http://www.thewabt.com/). 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 (www.uni-hamburg.de/IWI).

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 the followings: UCMSS (Universal Conference Management Systems & Support, California, USA) for managing all aspects of the conference; Dr. Tim Field of APC for managing the printing of the proceedings; and the staff of Monte Carlo Resort (Convention department) in Las Vegas for the professional service they provided. We would also like to thank Mahmoud Abou-Nasr for his continuous effort in organizing the special session on real-world data mining applications and serving as an Associate Co-Editor throughout the years. Additionally, we would like to thank the Co-Editors and Associate Co-Editors of DMIN'16: Diego Galar, and Peter Geczy.

Last but not least, we wish to express again our sincere gratitude and respect towards Prof. Hamid R. Arabnia (Univ. of Georgia, USA; Editor-in-Chief, Journal of Supercomputing (Springer)), General Chair and Coordinator of the federated congress, and also Associate Co-Editor of DMIN'16 for his excellent and tireless support, organization and coordination of all affiliated events. His exemplary and professional effort in 2016 and all the years before in the WORLDCOMP steering committee makes these events possible!

Thank you all for your contribution to DMIN'16! We hope that you will experience a stimulating conference with many opportunities for future contacts, research and applications.

We present the proceedings of DMIN'16.

Robert Stahlbock
DMIN'16 General Conference Chair

Gary M. Weiss

Steering Committee DMIN'16
www.dmin-2016.com

Contents

SESSION: REAL-WORLD DATA MINING APPLICATIONS, CHALLENGES, AND PERSPECTIVES

SESSION: DATA SCIENCE AND DATA SERVICES

SESSION: SEGMENTATION, CLUSTERING, ASSOCIATION + WEB / TEXT / MULTIMEDIA MINING + SOFTWARE

SESSION: REGRESSION AND CLASSIFICATION

SESSION: LATE POSTERS