

**PROCEEDINGS OF
THE 2014 INTERNATIONAL CONFERENCE ON
DATA MINING**

DMIN 2014

Editors

**Robert Stahlbock
Gary M. Weiss**

Associate Editors

**Mahmoud Abou-Nasr
Hamid R. Arabnia**



WORLDCOMP'14

July 21-24, 2014

Las Vegas Nevada, USA

www.world-academy-of-science.org

©CSREA Press

This volume contains papers presented at The 2014 International Conference on Data Mining (DMIN'14). Their inclusion in this publication does not necessarily constitute endorsements by editors or by the publisher.

Copyright and Reprint Permission

Copying without a fee is permitted provided that the copies are not made or distributed for direct commercial advantage, and credit to source is given. Abstracting is permitted with credit to the source. Please contact the publisher for other copying, reprint, or republication permission.

Copyright © 2014 CSREA Press
ISBN: 1-60132-267-4
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 10th International Conference on Data Mining 2014, DMIN'14 (www.dmin-2014.com), July 21-24, 2014, 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, and make better decisions. 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 big data, advanced- / business- / customer- / data- / ... / risk-analytics, to name only a few concepts that dominate websites, trade journals, and the general press. But they all aim at leveraging data for a better understanding of and insight into real-world phenomena. They all pursue this objective using some formal, often algorithmic, procedure, 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.

An important mission of the World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP, 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."* 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 78 different nations representing variety of personal and scientific experiences that arise from differences in culture and values. 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'14 assembles a spectrum of 21 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 21 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 2013 and the years before, we hope that the 2014 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'14 provides an international and multicultural experience with contributions from 18 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'14 attracted a large 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, taken 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 48%.

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'14 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'14 program committee members are (in alphabetical order): Mahmoud Abou-Nasr, Khalid Benabdeslem, Daniel Berrar, Jane Margot Binner, James Buckley, Peng Chen, Kevin Daimi, Christian Dawson, António Dourado, Mengling Feng, Philippe Fournier-Viger, Shunkai Fu, Peter Geczy, Corani Giorgio, Zahid Halim, Tzung-Pei Hong, Wei-Chiang Hong, Danish Irfan, Madjid Khalilian, Rikard König, Sebastian Klenk, Terje Kristensen, Philippe Lenca, Wen-Yang Lin, Qi Liu, Tanja Magoc, Jun Meng, José M. Merigo Lindahl, Sergey Morozov, Antonio Luigi Perrone, Hossein Peyvandi, Lotfi Ben Romdhane, Zhang Sen, Sabrina Senatore, Xuequn Shang, Victor Sheng, Yong Shi, Vijendra Singh, Robert Stahlbock, Baoying Wang, Chamont Wang, Simon Wang, Gary M. Weiss, Zijiang Yang, Faisal Zaman, and Yun Zhai.

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.), and Sven F. Crone (Lancaster University, UK) for circulating information on the conference.

Considering the increasing efforts of all towards the quality of the review process, the conference sessions and the social program of DMIN'14, 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: Diego Galar (Luleå University of Technology, Sweden), Peter Geczy (AIST, Japan), and Gary M. Weiss (Fordham University, USA).

The DMIN'14 conference organizers are also thankful to a number of co-sponsors, without whom the conference would not have been possible. As Sponsors-at-large, partners, and/or organizers each of the followings (separated by semicolons) provided help for at least one track of the World Congress: Computer Science Research, Education, and Applications Press (CSREA); Taylor & Francis, UK (<http://www.taylorandfrancis.com>); US Chapter of World Academy of Science (<http://www.world-academy-of-science.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 above and also on the cover of the proceedings), several publishers of computer science and computer engineering books and journals, chapters and/or task forces of computer science associations/organizations from 5 countries, 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) and the Business Intelligence Laboratory, B I³S lab, Hamburg, Germany (www.bis-lab.com).

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 throughout the years.

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); Editor-in-Chief, Emerging Trends in CS and Applied Computing (Elsevier); Editor-in-Chief, Transactions of Computational Science & Computational Intelligence (Springer); Elected Fellow, Int'l Soc. of Intelligent Biological Medicine (ISIBM); USA), General Chair and Coordinator of the federated congress, for his excellent and tireless support, organization and coordination of all affiliated events. His exemplary and professional effort in 2014 and all the years before in the WORLDCOMP steering committee makes these events possible!

Thank you all for your contribution to DMIN'14! We hope that you will experience a stimulating conference with many opportunities for future contacts, research and applications.

We present the proceedings of DMIN'14.

Robert Stahlbock
DMIN'14 General Conference Chair

Gary M. Weiss

Steering Committee DMIN'14
www.dmin-2014.com

Steering Committee WORLDCOMP, 2014
<http://www.world-academy-of-science.org/>

Contents

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

Mechanical Property Classification of Vapor-Grown Carbon Nanofiber/Vinyl Ester Nanocomposites Using Support Vector Machines	3
<i>Osama Abuomar, Sasan Nouranian, Roger King, Trenton Ricks, Thomas Lacy</i>	
APPES Maps as Tools for Quantifying Performance of Truck Drivers	10
<i>Iulian Carpatorea, Slawomir Nowaczyk, Thorsteinn Rognvaldsson, Marcus Elmer</i>	
Distributed Evolutionary Algorithm for Clustering Multi-Characteristic Social Networks	17
<i>Mustafa Hajeer, Dipankar Dasgupta, King-Ip Lin</i>	
Implementation of Artificial Neural Networks in MapReduce Optimization	26
<i>Changlong Li, XueHai Zhou, Kun Lu, Chao Wang, Dong Dai</i>	
A New Effective Information Decomposition Approach for Missing Data Recovery	30
<i>Shigang Liu, Honghua Dai</i>	
Predicting Causes of Traffic Road Accidents Using Multi-class Support Vector Machines	37
<i>Elfadil Mohamed</i>	
Gold, Oil and the S&P 500 Index: Calm to Crisis and Back	43
<i>Mary Malliaris, A.G. Malliaris</i>	
Using Bootstrap Aggregated Neural Networks for Peripheral Nerve Injury Treatment	49
<i>Munish Shah, Wei Chang, Kathleen McGuire, William Koch, Yan Meng, Xiaojun Yu</i>	
The Significance of the Race Factor in Breast Cancer Prognosis	55
<i>M. Mehdi Owrang-Ojaboni</i>	
A Proposed Data Mining Model for the Associated Factors of Alzheimer' s Disease	62
<i>Nevine Labib, Mohamed Badawy</i>	
Automated Statistical Data Mining of a Real World Landslide Detection System	69
<i>Divya Pullarkatt, Geethu Thottungal, Geethalekshmy V, Maneesha Vinodini Ramesh</i>	

SESSION: CLASSIFICATION, CLUSTERING, ASSOCIATION

Feature Selection by Tree Search of Correlation-Adjusted Class Distances	79
<i>Hassan Rabeti, Martin Burtscher</i>	

Multi-Level Synthesis of Frequent Rules from Different Databases Using a Clustering Approach	88
<i>Salim Khiat, HafiDa Belbachir, Sid Ahmed Rahal</i>	
Mammogram Classification Using Association Rule Mining	95
<i>Deepa Deshpande, Archana Rajurkar, Ramchandra Manthalkar</i>	
Feature Engineering for Supervised Link Prediction on Dynamic Social Networks	102
<i>Jeyanthi Narasimhan, Lawrence Holder</i>	
A Flexible Feature Selection Framework for Improving Breast Cancer Classification from Sparse Spectral Count Proteomic Data	109
<i>Lanfei Shi, Himanshu Grover, Jeya Balaji Balasubramanian, Kumar Kolli, Craig Shriver, Vanathi Gopalakrishnan</i>	
Optimization of an Individual Re-identification Modeling Process Using Biometric Features	116
<i>Alejandro Heredia-Langner, Brett Amidan, Shari Matzner, Kristin Jarman</i>	
More Reliable Over-sampled Synthetic Data Instances by Using Artificial Neural Networks for a Minority Class	121
<i>Hyontai Sug, Douglas Dankel</i>	
Entropy Based Adaptive Outlier Detection Technique for Data Streams	125
<i>Yogita Thakran, Durga Toshniwal, Bhavani Kumar Eshwar</i>	
SESSION: WEB, TEXT, MULTIMEDIA MINING	
Sub-Net Identification for Text Refinement and Text Meaning Discovery in Social Media Analytics	135
<i>Dinesh Batra, Gurpreet Bawa, Vivek Anand</i>	
Semi-automatic Metadata Extraction from Scientific Journal Article for Full-text XML Conversion	142
<i>Sukyong Kim, Yoonsung Cho, Kihong Ahn</i>	
The Authorship of Audacity: Data Mining and Stylometric Analysis of Barack Obama Speeches	148
<i>Jonathan Herz, Abdelghani Bellaachia</i>	
Text Ctegorization using Topic Model and Ontology Networks	155
<i>Yinghao Huang, Xipeng Wang, Yi Murphey</i>	

Personalizing Query Refinement Based on Latent Tasks 162
Chao Xu, Mingzhu Zhu, Yanchi Liu, Yi-Fang Wu

Positive Unlabeled Learning to Discover Relevant Documents Using Topic Models for Feature Selection 169
Mingzhu Zhu, Chao Xu, Yi-Fang Wu

A Framework For Flexible Educational Data Mining 176
Kyle DeFreitas, Margaret Bernard

SESSION: LATE BREAKING PAPERS AND POSITION PAPERS: DATA MINING

Author Attribution of Thomas Paine Work 183
Smiljana Petrovic, Gary Berton, Robert Schiaffino, Lubomir Ivanov

Identifying Outliers in Human Movement Trajectories Clustered By Hausdorff Distance 190
Shibin Parameswaran, Jeffrey Ellen

Automatic Solar Cavity Detection Using Haar Cascade Classifier 197
Aijuan Dong, Athena Johnson, George Dimitoglou, Joshua Shelley

Processing of Kuala Lumpur Stock Exchange Resident on Hadoop MapReduce 204
Hui Kim Law, Saeed Aghabozorgi, Say Er Lim, Ying Wah Teh , Tutut Herawan

Generating Well-Behaved Learning Curves: An Empirical Study 210
Gary Weiss, Alexander Battistin

Determination of Fake Reviews in Hospitality Sector 214
Manish Goswami, S K Gupta

