PROCEEDINGS OF
THE 2013 INTERNATIONAL CONFERENCE ON
DATA MINING

DMIN 2013

Editors

Robert Stahlbock
Gary M. Weiss

Associate Editors

Mahmoud Abou-Nasr
Hamid R. Arabnia

WORLDCOMP'13
July 22-25, 2013
Las Vegas Nevada, USA
www.world-academy-of-science.org

©CSREA Press

This volume contains papers presented at The 2013 International Conference on Data Mining (DMIN'13). 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 © 2013 CSREA Press
ISBN: 1-60132-239-9
Printed in the United States of America

CSREA Press
U. S. A.

Foreword

We are pleased to present this collection of papers submitted to the 9th International Conference on Data Mining 2013, DMIN'13 (www.dmin-2013.com), July 22-25, 2013, The New Tropicana Hotel, 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. 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, the largest annual gathering of researchers in computer science, computer engineering and applied computing. WORLDCOMP'13 assembles a spectrum of 22 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 22 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 2012 and the years before, we hope that the 2013 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. The 2013 conference has provided an international and multicultural experience with contributions from 26 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'13 attracted a large number of submissions of theoretical research papers as well as industrial reports and case studies on applications. The program committee would like to thank all those who submitted papers for review. 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 45%.

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 DMIN'13 program committee. Their continuing support has been essential to further improve the quality of accepted submissions and the resulting success of the conference. The DMIN'13 program committee members are (in alphabetical order):

Mahmoud Abou-Nasr, Lamine M. Aouad, Jérôme Azé, Souhaib Ben Taieb, Daniel Berrar, James Buckley, Alina Campan, Peng Chen, Paulo Cortez, Christian Dawson, Qin Ding, António Dourado, Mengling Feng, Philippe Fournier-Viger, Shunkai Fu, Peter Géczy (JP), Corani Giorgio, Zahid Halim, Tzung-Pei Hong, Yo-Ping Huang, Ulf Johansson, Rikard König, Madjid Khalilian, Sebastian Klenk, Terje Kristensen, Philippe Lenca, Chuan Li, Wen-Yang Lin, José M. Merigo Lindahl, Tanja Magoc, Jun Meng, Sergey Morozov, Hossein Peyvandi, Torsten Reiners, Lotfi Ben Romdhane, Gerald Schaefer, Sabrina Senatore, Xuequn Shang, Victor Sheng, Yong Shi, Vijendra Singh, Robert Stahlbock, Shun-Hung Tsai, Nicole Vincent, Baoying Wang,

Chamont Wang, Fan Wang, Simon Wang, Xuewei Wang, Gary M. Weiss, Zijiang Yang, Yun Zhai, Defu Zhang, and Shang-Ming Zhou.

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., Intelligent Systems Instructor, Trailblazer Intelligent Systems, 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'13 we are confident that you can look forward to participating and attending a leading and reputable international conference. 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: Vladimir Cherkassky (University of Minnesota, USA), Peter Geczy (AIST, Japan), and Alfred Inselberg (Tel Aviv University, Israel).

The DMIN'13 conference organizers are also thankful to a number of co-sponsors, without whom the conference would not have been possible. The conference was sponsored and or supported by the US Chapter of World Academy of Science (http://www.world-academy-of-science.org/); Computer Science Research, Education, and Applications Press (CSREA); Aldebaran Robotics Inc., USA; Taylor & Francis, United Kingdom. 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 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 I^{3}S lab, Hamburg, Germany (www.bis-lab.com). Furthermore, we would like to thank all members of the steering committee of the 2013 congress: Dr. Selim Aissi, (formerly: Chief Strategist - Security, Intel Corporation, USA), Senior Business Leader & Head of Global Enterprise Security Architecture, Visa Corporation, USA; Prof. Nizar Al-Holou, Professor and Chair, Electrical and Computer Engineering Department, Vice Chair, IEEE/SEM-Computer Chapter, University of Detroit Mercy, Detroit, Michigan, USA; Prof. Hamid R. Arabnia, Professor of Computer Science, Editor-in-Chief of The Journal of Supercomputing (Springer), Elected Fellow of Int'l Society of Intelligent Biological Medicine (ISIBM), The University of Georgia, Georgia, USA; Prof. Babak Akhgar, Fellow of the British Computer Society, CITP, Professor of Informatics, Co-Director of CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence & organised Crime research), Sheffield Hallam University, Sheffield, UK; Prof. Kevin Daimi, Director, Computer Science and Software Engineering Programs, Department of Mathematics, Computer Science and Software Engineering, University of Detroit Mercy, Detroit, Michigan, USA; Prof. George Jandieri, Chief Scientist, The Institute of Cybernetics, Georgian Academy of Science, Georgia, Editorial Board Member of International Journal of Microwaves and Optical Technology, The Open Atmospheric Science Journal, American Journal of Remote Sensing, Georgian Technical University, Tbilisi, Georgia; Prof. Dattatraya V. Kodavade, Head of Computer Science and Engineering, D.K.T.E Society's Textile & Engineering Institute, Ichalkaranji, Maharashtra State, India; Prof. George Markowsky, Professor and Associate Director, School of Computing and Information Science, Chair of International Advisory Board of IEEE IDAACS, Director 2013 Northeast Collegiate Cyber Defense Competition, Chair Bangor Foreign Policy Forum, President Phi Beta Kappa Delta Chapter of Maine, Cooperating Professor Mathematics and Statistics Department UMaine, Cooperating Professor School of Policy & International Affairs UMaine, University of Maine, Orono, Maine, USA; Prof. G. N. Pandey, Vice-Chancellor, Arunachal University of Studies, Arunachal Pradesh, India and Adjunct Professor, Indian Institute of Information Technology, Allahabad, India; Prof. Fernando G. Tinetti, Co-editor of Journal of Computer Science and

Technology (JCS&T), Professor, School of Computer Science, Universidad Nacional de La Plata, La Plata, Argentina; Dr. Predrag Tosic, Microsoft, Washington, USA; Prof. Vladimir Volkov, The Bonch-Bruevich State University of Telecommunications, Saint-Petersburg, Russia.

Most importantly, we wish to express again our sincere gratitude and respect towards Professor Hamid R. Arabnia (Univ. of Georgia, USA; Elected Fellow, ISIBM; Editor-in-Chief, The Journal of Supercomputing (Springer); Member), 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 2013 and all the years before in the WORLDCOMP steering committee makes these events possible!

Thank you all for your contribution to DMIN'13! We hope that you will experience a stimulating conference with many opportunities for future contacts, research and applications.

Robert Stahlbock
DMIN'13 General Conference Chair

Gary M. Weiss

Steering Committee DMIN'13
www.dmin-2013.com

Contents

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

Maintenance Knowledge Management with Fusion of CMMS and CM **3**
Sten-Erik Bjorling, Diego Galar, David Baglee, Sarbjeet Singh, Uday Kumar

Sentimental Analysis on Turkish Blogs via Ensemble Classifier **10**
Sadi Evren Seker, Khaled Al -Naami

Reliable Probabilistic Classification of Mammographic Masses Using Random Forests **16**
Hechmi Shili, Lotfi Ben Romdhane, Bechir El Ayeb

Identifying Patterns and Anomalies in Delayed Neutron Monitor Data of Nuclear Power Plant **23**
Durga Toshniwal, Aditya Gupta, Pramod Gupta, Vikas Khurana, Pushp Upadhyay

Alleviating the Class Imbalance Problem in Data Mining **30**
Akkenzhe Sarmanova, Songul Albayrak

Efficiency of Crop Yield Forecasting Depending on the Moment of Prediction based onLarge Remote Sensing Data set **36**
Alexander Murynin, Konstantin Gorokhovskiy, Vladimir Ignatiev

Neural Network Forecasting with the S&P 500 Index Across Decades **41**
Mary Malliaris, Anastasios G. Malliaris

Data Uncertainty Handling in High Level Information Fusion **47**
Robert Woodley, Michael Gosnell, Amber Fischer

A Preliminary Approach to Study the Causality of Freezing of Gait for Parkinson's: Bayesian Belief Network Approach **53**
Ali Saad, Abass Zeineldine, Iyad Zaarour, Mohammad Ayache, Dimitri Lefebvre, Francois Guerin, Paul Bejjani

Evaluation of Monte Carlo Subspace Clustering with OpenSubspace **58**
David Hunn, Clark Olson

MineTool-3DM2: An Algorithm for Data Mining of 3D Simulation Data **65**
Tamara Sipes, Homa Karimabadi

Actions Ontology System for Action Rules Discovery in Mammographic Mass Data **73**
Angelina A. Tzacheva, Erik A. Koenig, Justin R. Pardue

GDP Forecasting through Data Mining of Seaport Export-Import Records **79**
H. Raymond Joseph

Association Rule Mining for Finding Correlations Among People **83**
Valmik Nikam, Nimai Buch, Yash Botadra, B.B. Meshram

Toward Sustainable High-Yield Agriculture via Intelligent Control Systems **88**
Brian McLaughlan, James Brandli

Extending Local Similarity Indexes with KNN for Link Prediction **92**
Greg Speegle, Yilan Bai, Young-Rae Cho

A New Simple Classification Algorithm enabling a New Approach for Identification of Virtual Bullying **99**
K Burn-Thornton, T Burman

Using Data Mining to Analyze Donation Data for a Local Food Bank **105**
Steven Jiang, Lauren Davis, Henrique Tavares De Mleo, Jessica Terry

Flash Reactivity: Adaptative models in recommender systems **111**
Julien Gaillard, Marc El-Beze, Eitan Altman, Emmanuel Ethis

Analysis of Truck Compressor Failures Based on Logged Vehicle Data **118**
Rune Prytz, Sławomir Nowaczyk, Thorsteinn Rognvaldsson, Stefan Byttner

Proposed Business Intelligence Models for Medical Risk Assessment Case study of Venous Thrombosis Disease in Egypt **125**
Edward Wadid, Nevine Makram Labib, Sayed Abdel Wahab

Improve the Quality of Product Recommendation based on Multi-channel CRM for E-commerce **132**
Chuen-He Liou

Using Recursive Sorting to Improve Accuracy of Memory-based Collaborative Filtering Recommendations **138**
Serhiy Morozov, Hossein Saiedian

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

Mining for Hydrologic Features in LiDAR Data **145**
Rebecca Reizner, Eric Shaffer, Brianna Birman

Role of Social Media in Early Warning of Norovirus Outbreaks: A Longitudinal Twitter-Based Infoveillance **149**
Ahmed YoussefAgha, Wasantha Jayawardene, David Lohrmann

Spatial-Temporal Clustering of a Self-Organizing Map **154**
Carlos Enrique Gutierrez, Mohammad Reza Alsharif, Katsumi Yamashita, Rafael Villa, Cuiwei He, Hayao Miyagi

An Evolutionary Associative Contrast Rule Mining Method for Incomplete Database **160**
Kaoru Shimada, Takashi Hanioka

Hierarchical Video Indexing And Retrieval System **167**
Mohammed Yassine Kazi Tani, Abdelghani Ghomari, Lamia Dali Youcef

A Novel Query Suggestion Method Based On Sequence Similarity and Transition Probability **174**
Bo Shu, Zhendong Niu, Xiaotian Jiang, Ghulam Mustafa

SESSION: REGRESSION AND CLASSIFICATION

A Multi-scale Nonparametric/Parametric Hybrid Recognition Strategy with Multi-category Posterior Probability Estimation **183**
Zhao Lu, Zheng Lu, Haoda Fu

SVM-Based Approaches for Predictive Modeling of Survival Data **190**
Han-Tai Shiao, Vladimir Cherkassky

Large Scale Visual Classification with Parallel, Imbalanced Bagging of Incremental LIBLINEAR SVM **197**
Thanh-Nghi Doan, Thanh-Nghi Do, Francois Poulet

Gaussian Process Regression with Dynamic Active Set and Its Application to Anomaly Detection **204**
Toshikazu Wada, Yuki Matsumura, Shunji Maeda, Hisae Shibuya

A Study of kNN using ICU Multivariate Time Series Data **211**
Dan Li, Admir Djulovic, Jianfeng Xu

Genetic Algorithms and Classification Trees in Feature Discovery: Diabetes and the NHANES database **218**
Alejandro Heredia-Langner, Kristin Jarman, Brett Amidan, Joel Pounds

SESSION: FILTERING, FEATURE SELECTION, INTEGRATION, ENSEMBLES

Isolating Matrix Sparsity in Collaborative Filtering Ratings Matrices **225**
Brian Pechkis, Eun-Joo Lee

Brian Pechkis, Eun-Joo Lee

A Novel Randomized Feature Selection Algorithm **231**
Subrata Saha, Rampi Ramprasad, Sanguthevar Rajasekaran

Fraud Detection Using Reputation Features, SVMs, and Random Forests **238**
Dave DeBarr, Harry Wechsler

A Novel Ensemble Selection Technique for Weak Classifiers **245**
Kung-Hua Chang, D. Stott Parker

Labeled Subgraph Matching Using Degree Filtering **252**
Lixin Fu, Surya Kommireddy

Social Network Anonymization and Influence Preservation **258**
Alina Campan, Yasmeen Alufaisan