

**PROCEEDINGS OF
THE 2012 INTERNATIONAL CONFERENCE ON
DATA MINING**

DMIN²⁰¹²

Editors

**Robert Stahlbock
Gary M. Weiss**

Associate Editors

Mahmoud Abou-Nasr, Hamid R. Arabnia



WORLD COMP'12

July 16-19, 2012

Las Vegas Nevada, USA

www.world-academy-of-science.org

©CSREA Press

This volume contains papers presented at The 2012 International Conference on Data Mining (DMIN'12). 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 © 2012 CSREA Press
ISBN: 1-60132-208-9
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 8th International Conference on Data Mining 2012, DMIN'12 (www.dmin-2012.com), July 16-19, 2012, 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. DMIN publishes and promotes this work with contributions that focus on research and practice from a wide number of application areas.

DMIN conferences are held annually within the federated WORLDCOMP congress, the largest annual gathering of researchers in computer science, computer engineering and applied computing. An important mission of WORLDCOMP is to provide 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. WORLDCOMP'12 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 2011 and the years before, we hope that the 2012 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 conferences actively support 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 2012 conference has provided an international and multicultural experience with contributions from 25 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'12 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 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. Each paper was refereed by at least two researchers in the topical area, with most of the papers receiving three reviews, and the reviewer's expertise and confidence were considered when rendering the final decision. The review process was highly competitive. The overall paper acceptance rate for papers was 44%

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'12 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'12 program committee members are (in alphabetical order): Mahmoud Abou-Nasr, Jérôme Azé, Souhaib Ben Taieb, Khalid Benabdeslem, James Buckley, Alina Campan, Pedro A. Castillo, Peng Chen, Paulo Cortez, Kevin Daimi, António Dourado, Mohammed Farquad, Mengling Feng, Philippe Fournier-Viger, Shunkai Fu, Peter Géczy, Corani Giorgio, Zahid Halim, Tran Hoai Linh, Tzung-Pei Hong, Wei-

Chiang Hong, Mehrdad Jalali, Nathalie Japkowicz, Madjid Khalilian, Sebastian Klenk, Terje Kristensen, Abhijit Kulkarni, Jaewook Lee, Chuan Li, Wen-Yang Lin, Jing Liu, Weifeng Liu, Antonio Luigi Perrone, Lamine M. Aouad, Tanja Magoc, Sergey Morozov, Hossein Peyvandi, Mekki Rachida, Torsten Reiners, Gerald Schaefer, Zhang Sen, Sabrina Senatore, Victor Sheng, Yong Shi, Tamanna Siddiqui, Vijendra Singh, Robert Stahlbock, Shiliang Sun, Sundaram Suresh, Jaakko Talonen, Shun-Hung Tsai, Nicole Vincent, Baoying Wang, Chamont Wang, Fan Wang, Simon Wang, Xuwei Wang, Gary Weiss, Zhixian Yan, Zijiang Yang, Faisal Zaman, Yun Zhai, Defu Zhang, Songfeng Zheng, 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, as well as Gregory Piatetsky-Shapiro (editor KDNuggets), for publishing DMIN announcements.

Considering the increasing efforts of all towards the quality of the review process, the conference sessions, and the social program, we are confident of the success of DMIN'12. It is a particular pleasure to provide data mining oriented invited talks by the following esteemed members of the data mining community: Sofus A. Macskassy (Univ. of Southern California, USA), Peter Geczy (AIST, Japan), and Haym Hirsh (Rutgers University, USA).

The DMIN'12 conference organizers would also like to thank co-sponsors, without whom the conference would not have been possible. The co-sponsors of this year's WORLDCOMP conferences include (separated by semicolons): Bioinformatics & Computational Biology Program, George Mason University, Virginia, USA (<http://www.binf.gmu.edu/index.html>); Biomedical Cybernetics Laboratory, HST of Harvard University and MIT, USA (<http://bcl.med.harvard.edu/>); Minnesota Supercomputing Institute, University of Minnesota, USA (<http://www.msi.umn.edu/>); Center for Cyber Defense, NCAT (<http://caei.ncat.edu/CCD/>); Argonne's Leadership Computing Facility of Argonne National Laboratory, Illinois, USA (<http://www.alcf.anl.gov/>); The Center for Advanced Studies in Identity Sciences (CASIS: NC A&T, Carnegie Mellon, Clemson, UNC Wilmington - <http://casis.ncat.edu/>); Knowledge Management & Intelligent System Center (KMIS) of University of Siegen, Germany (<http://www.kmis.uni-siegen.de/>); Intelligent Cyberspace Engineering Lab., ICEL, Texas A&M University, Commerce, Texas, USA; UMIT, Institute of Bioinformatics and Translational Research, Austria; Hawkeye Radiology Informatics, Department of Radiology, College of Medicine, University of Iowa, Iowa, USA (<http://www.uiowa.edu/~hri/>); The International Council on Medical and Care Compunetics (<http://www.icmcc.org>); US Chapter of World Academy of Science (<http://www.world-academy-of-science.org/>); Supercomputer Software Department (SSD), Institute of Computational Mathematics & Mathematical Geophysics, Russian Academy of Sciences (<http://ssd.ssc.ru>); International Society of Intelligent Biological Medicine (<http://www.isibm.org/>); NDSU-CIIT Green Computing and Communications Laboratory (<http://gcclab.org/>); Medical Image HPC & Informatics Lab (MiHi Lab), University of Iowa, Iowa, USA (<http://www.uiowa.edu/mihpclub/>); High Performance Computing for Nanotechnology (HPCNano - <http://www.hpcnano.org>); Manx Telecom (<http://www.manxtelecom.com/>); Computer Science Research, Education, and Applications Press; World Academy of Biomedical Sciences and Technologies (<http://www.worldwabt.org/wabt>); HoIP Telecom (<http://www.hoip-telecom.com>); Super Micro Computer, Inc., San Jose, California, USA (<http://www.supermicro.com>); Intel Corporation (<http://www.intel.com>); Hodges Health (<http://www.p-jones.demon.co.uk/>); and OMG™ (<http://www.omg.org/>). In addition, a number of university faculty members and their staff, several publishers of computer science and computer engineering books and journals, chapters and/or task forces of computer science associations/organizations from 6 countries, and developers of high-performance machines and systems provided significant help in organizing the WORLDCOMP conferences 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), CSREA Computer Science Research, Education, and Applications Press, and the Business Intelligence Laboratory, B I³S lab, Hamburg, Germany (www.bis-lab.com).

Furthermore, we want to thank all members of the steering committee of WORLDCOMP 2012: Prof. Babak Akhgar, PhD, FBCS, CITP, Professor of Informatics, Sheffield Hallam University, Sheffield, UK; Prof. Gerry Vernon Dozier, Chair, Department of Computer Science; Director, Center for Advanced Studies in Identity Sciences; Center for Cyber Defense; North Carolina A&T State University, North Carolina, USA; Prof. Madjid Fathi, Director, Knowledge Management and Intelligent Systems Center, University of Siegen, Germany; Prof. Michael R. Grimaila, Air Force Institute of Technology, Systems Engineering; Fellow of ISSA; CISM, CISSP, IAM/IEM; Editorial Board of ISSA Journal; Air Force Center of Cyberspace Research; Advisor to the Prince of Wales Fellows & Prince Edward Fellows at MIT and Harvard Universities; PC member, NATO Cooperative Cyber Defence Centre of Excellence (CCD COE) & Int'l Conf. on Information Warfare and Security; Prof. Kun Chang Lee, Professor of MIS and WCU Professor of Creativity Science, Business School and Department of Interaction Science, Sungkyunkwan University, Seoul, South Korea; Prof. Dr. Dr.h. Victor Malyshev, Head, Supercomputer Software Department (SSD), Institute of Computational Mathematics and Mathematical Geophysics, Russian Academy of Sciences, Russia; Prof. George Markowsky, Associate Director, School of Computing and Information Science; Chair International Advisory Board of IEEE IDAACS; Director 2013 Northeast Collegiate Cyber Defense Competition; Chair Bangor Foreign Policy Forum; Cooperating Professor Mathematics and Statistics Department UMaine; Cooperating Professor School of Policy & International Affairs UMaine; University of Maine, Orono, Maine, USA; Prof. Andy Marsh, Director HoIP; Director HoIP Telecom, UK; Secretary-General WABT; Vice-president ICET; Visiting Professor University of Westminster, UK; Prof. James J. (Jong Hyuk) Park, Department of Computer Science and Engineering, Seoul National University of Science and Technology (SeoulTech), Korea; President, KITCS; President, FTRA; Editor-in-Chiefs: HCIS, JoC and IJITCC Journals; Dr. Hassan Reza, UND Aerospace, University of North Dakota, Department of Computer Science, Grand Forks, North Dakota, USA; Prof. Sang C. Suh, Head and Professor, Department of Computer Science; Vice President, Society for Design and Process Science (SDPS); Director, Intelligent Cyberspace Engineering Lab (ICEL); Texas A&M University, Commerce, Texas, USA; Prof. Layne T. Watson, IEEE Fellow; NIA Fellow; ISIBM Fellow; Fellow of The National Institute of Aerospace; Virginia Polytechnic Institute & State University, Virginia, USA.

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, Advisory Board, IEEE TC on Scalable Computing), General Chair and Coordinator of all WORLDCOMP conferences, for his excellent and tireless support, organization and coordination of all affiliated events. His exemplary and professional effort in 2012 and all the years before in the WORLDCOMP steering committee makes these events possible.

Thank you all for your contribution to DMIN'12. We hope that you will experience a stimulating conference with many opportunities for future contacts, research and applications.

Robert Stahlbock
Gary M. Weiss

Steering Committee DMIN'12
www.dmin-2012.com

Contents

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

Results of Mining Data Features During Computational Fluid Dynamics Simulations	3
<i>Michael R. Gosnell, Robert S. Woodley, Steven E. Gorrell</i>	
Prediction of Pull-out capacity of Suction Caissons Using Self-Evolving Neural Networks	10
<i>Abdussamad Ismail, Dong-Sheng Jeng</i>	
Model for Aggregated Water Heater Load Using Dynamic Bayesian Networks	17
<i>Maria Vlachopoulou, George Chin, Jasson C Fuller, Shuai Lu, Karanjit Kalsi</i>	
Integrating Decision Tree and K-Means Clustering with Different Initial Centroid Selection Methods in the Diagnosis of Heart Disease Patients	24
<i>Mai Shouman, Tim Turner, Rob Stocker</i>	
Forecasting Stock Price Movement with Semi-Supervised Learning	31
<i>Kanghee Park, Hyunjung Shin</i>	
Finding Interesting Classification Rules: An Application from Education	37
<i>Anthony Scime, Summer M. Reiner</i>	
Using Random Probes for Neural Networks Based Features Selection	44
<i>Hazem Migdady, Norman Carver</i>	
Fraudulent Bill-Claim Detection in Health Insurance	50
<i>Junwoo Lee, Juhyeon Kim, Hyunjung Shin</i>	
How a Financial Crisis Affects Data Mining Results: A Case Study	57
<i>Mary Malliaris, Anastasios G. Malliaris</i>	
Constrained Nonnegative Matrix Factorization based Feature Selection	63
<i>Nirmal Thapa, Jun Zhang</i>	
Generating Rules to Increase Production Using Decision Tree	69
<i>Keivan Ghoseiri, Hassan Gholami Mazinan, Mahyar Hoseinzadeh, Maziar Davoodi, Erfan Khaji</i>	
Classification and Regression Trees for Handling Missing Values in a CMBD to reduce malware in an Information System	73
<i>Gustavo A Valencia-Zapata, Juan C Salazar-Uribe</i>	

SPARCL: An Improved Approach for Matching Sinhalese Words and Names in Record Clustering and Linkage	79
<i>Gayan Prasad Hettiarachchi, Dihari Attygalle</i>	

The Importance of Tuning Financial Technical Indicators to Predict Stock Movements	85
<i>Maysa Ammouri, Sameh Al-Shihabi</i>	

SESSION: SEGMENTATION, CLUSTERING, ASSOCIATION

Cartogram Data Projection for Self-Organizing Maps	91
<i>David Brown, Lutz Hamel</i>	

A Population Based Convergence Criterion for Self-Organizing Maps	98
<i>Lutz Hamel, Benjamin Ott</i>	

Clustering Approaches for Financial Data Analysis: a Survey	105
<i>Fan Cai, Nhien-An Le-Khac, M-Tahar Kechadi</i>	

Consensus Clustering from Experts' Partitions for Patients' nevi: Model the Ugly Duckling	112
<i>Yanal Wazaefi, Yvane Bruneu, Julien Lefevre, Gloria Menegaz, Giulia Paggetti, Arnaud Le Troter, Sebastien Paris, Jean-Jacques Grob, Bernard Fertil</i>	

Intrusion Detection System with Data Stream Clustering Approach	119
<i>Madjid Khalilian, Norwati Mustapha, Nasir Suliman, Ali Mamat</i>	

Towards Processing Multi-Dimensional Dynamic Data	126
<i>Yong Shi, Brian Graham</i>	

SESSION: REGRESSION AND CLASSIFICATION + FEATURE SELECTION

MineTool-M2: An Algorithm for Data Mining of 2D Simulation Data	133
<i>Tamara Sipes, Homa Karimabadi</i>	

Hybrid Predictive Models for Optimizing Marketing Banner Ad Campaign in On-line Social Network	140
<i>Mariusz Lapczynski, Jerzy Surma</i>	

Application of Data Mining Techniques to Predict Allergy Outbreaks among Elementary School Children: Integration of Hourly Air Pollution, Bi-Daily Upper-Air, and Daily School Health Surveillance Systems in Pennsylvania	147
<i>Ahmed YoussefAgha, Wasantha Jayawardene, David Lohrmann, Gamal El Afandi</i>	

Optimization and Evaluation Criteria in GP Regression	154
<i>Rikard Konig, Ulf Johansson, Lars Niklasson</i>	

Automated Provisioning of Campaigns Using Data Mining Techniques	161
<i>Saravanan M, Deepika S.M.</i>	

The Impact of the Observation of Predictive Features on The Diagnosis of Pigmented Skin Lesions and the Therapeutic Decision	168
<i>Yanal Wazaefi, Arthur Tenenhaus, Alex Nkengne, Jean-François Horn, Alain Giron, Sebastien Paris, Bernard Fertil</i>	

Conceptualization of Sentence Paraphrase Recognition with Semantic Role Labels	172
<i>Ritu Yadav, Ashwani Kumar, A. Vinay Kumar, Pradeep Kumar</i>	

SESSION: WEB AND TEXT MINING

Analyzing Conflict Narratives to Predict Settlements in EBay Feedback Dispute Resolution	181
<i>Xiaoxi Xu, David A. Smith, Tom Murray, Beverly Park Woolf</i>	

Bayesian Model Averaging of Named Entity Extraction Algorithms	188
<i>Paul Kidwell, Kofi Boakye, Joseph Guensche, Ron Glaser, William Hanley, Tracy Lemmond</i>	

Location-Based Burst Detection Algorithm in Spatiotemporal Document Stream	195
<i>Keiichi Tamura, Hajime Kitakami</i>	

Definition of Table Similarity for News Article Classification	202
<i>Taeho Jo</i>	

Early Results of Composite NER Algorithm for Resume Corpora Distillation	208
<i>Sahil Patwardhan, Pallavi Agarwal, Pooja Sunder</i>	

A Content Analysis of Online News Media Reporting on American Health Care Reform	215
<i>Ahmed YoussefAgha, Wasantha Jayawardene, Samuel Obeng, David Lohrmann</i>	

Automatic Construction of Similarity Matrix for Semantic Numerical Operations on Strings	222
<i>Taeho Jo</i>	

Mining Social Data with UCL's SocialSTORM Platform	228
<i>Reece Wood, Ilya Zheludev, Philip Treleaven</i>	

SESSION: NOVEL APPLICATIONS AND ALGORITHMS

A Comparative Study of Predicting User Preference using Evolutionary Clustering Algorithm	235
<i>Chhavi Rana, Sanjay Kumar Jain</i>	

Parallelization Strategies for Distributed Non Negative Matrix Factorization	242
<i>Ahmed Nagy, Massimo Coppola, Nicola Tonellotto</i>	
Filtering Search Results using Explicit Feedback	249
<i>Varun Gupta, Neeraj Garg, Kapil Jhamb, Lakshya Bhagat</i>	
Credo: A Framework for Semi-supervised Credibility Assessment for Social Networks	257
<i>Ahmed Nagy, Jeannie Stamberger</i>	
Automatic Multi-label Categorization of News Feeds	264
<i>Majid Darabi, Hossein Adeli, Nasseh Tabrizi</i>	