

FOREWORD

These proceedings contain the papers of the IADIS European Conference on Data Mining 2007, which was organised by the International Association for Development of the Information Society in Lisbon, Portugal, July 5 – 7, 2007. This year, this conference is part of the Multi Conference on Computer Science and Information Systems 2007, 3 - 8 July 2007, which had a total of 1091 submissions.

The First European Conference on Data Mining (ECDM'07) is aimed to gather researchers and application developers from a wide range of data mining related areas such as statistics, computational intelligence, pattern recognition, databases and visualization. ECDM'07 is aimed to advance the state of the art in data mining field and its various real world applications. ECDM'07 will provide opportunities for technical collaboration among data mining and machine learning researchers around the globe.

Submissions were accepted under the following areas and topics:

→ Core Data Mining Topics

- Parallel and distributed data mining algorithms
- Data streams mining
- Graph mining
- Spatial data mining
- Text video, multimedia data mining
- Web mining
- Pre-processing techniques
- Visualization
- Security and information hiding in data mining

→ Data Mining Applications

- Databases
- Bioinformatics
- Biometrics
- Image analysis
- Financial modeling
- Forecasting
- Classification
- Clustering

The IADIS European Conference on Data Mining 2007 received 81 submissions from more than 24 countries. Each submission has been anonymously reviewed by an average of four independent reviewers, to ensure that accepted submissions were of a high standard. Consequently only 15 full papers were approved which means an acceptance rate below 19 %. A few more papers were accepted as short papers, reflection papers and posters. The best paper authors will be invited to publish extended versions of their papers in the IADIS International Journal on Computer Science and Information Systems (ISSN: 1646-3692) and also in the following Journals:

- International Journal of Computational Intelligence Research (special issue on advances in data mining);
- Journal of Information Assurance and Security (special issue on data mining and information security).

Besides the presentation of full papers, short papers, reflection papers and posters, the conference also included two keynote presentations from internationally distinguished researchers. We would therefore like to express our gratitude to Professor Janusz Kacprzyk, Deputy Director for Scientific Affairs, Polish Academy of Sciences, Poland, and to Professor Bogdan Gabrys, Computational Intelligence Research Group, Bournemouth University, UK, for accepting our invitation as keynote speakers. We wish also to thank the tutorial by Professor Janusz Kacprzyk, Deputy Director for Scientific Affairs, Polish Academy of Sciences, Poland.

As we all know, organising a conference requires the effort of many individuals. We would like to thank all members of the Program Committee, for their hard work in reviewing and selecting the papers that appear in the proceedings.

This volume has taken shape as a result of the contributions from a number of individuals. We are grateful to all authors who have submitted their papers to enrich the conference proceedings. We wish to thank all members of the organizing committee, delegates, invitees and guests whose contribution and involvement are crucial for the success of the conference.

Last but not the least, we hope that everybody will have a good time in Lisbon, and we invite all participants for the next year edition of the IADIS European Conference on Data Mining 2008.

Ajith Abraham,
Norwegian University of Science and Technology,
Norway
IADIS European Conference on Data Mining 2007 Conference Program Chair

Piet Kommers, University of Twente, The Netherlands
Pedro Isaías, Universidade Aberta (Portuguese Open University), Portugal
Nian-Shing Chen, National Sun Yat-sen University, Taiwan
MCCSIS 2007 General Conference Co-Chairs

Lisbon, Portugal
July 2007

PART III

IADIS EUROPEAN CONFERENCE

DATA MINING 2007

TABLE OF CONTENTS

FOREWORD	vii
PROGRAM COMMITTEE	ix
KEYNOTE LECTURES	xiii
CONFERENCE TUTORIAL	xvi

FULL PAPERS

PERSONALIZING TEXT SUMMARIZATION BASED ON SENTENCE WEIGHTING <i>Christos Bouras, Vassilis Pouloupoulos and Vassilis Tsogkas</i>	3
A DEMPSTER-SHAFER BELIEF THEORY BASED ALGORITHM FOR CLASSIFICATION OF IMPERFECT DATA <i>Jinsong Zhang, Lakshitha Weerakkody and David Roelant</i>	11
ON THE SUITABILITY OF TIME SERIES ANALYSIS ON DATA WAREHOUSES <i>Jose Zubcoff, Juan Trujillo and Alfredo Cuzzocrea</i>	17
EVALUATING A LEARNING ALGORITHM SELECTION METHOD FOR A RULE EVALUATION SUPPORT TOOL BASED ON OBJECTIVE INDICES <i>Hidenao Abe, Shusaku Tsumoto, Miho Ohsaki and Takahira Yamaguchi</i>	25
TOWARD A SEMANTIC DATA RETRIEVAL IN DISTRUIBUTED RELATIONAL DATABASE <i>A. Chella, M. Liotta, N. Ingraffia, G. Pilato and G. Vassallo</i>	33
ON APPLICATION OF PROBABILITY METRICS IN THE CLUSTER STABILITY PROBLEM <i>Z. Volkovich and Z. Barzily</i>	41
A WEB RECOMMENDER SYSTEM BASED ON ANT COLONY OPTIMIZATION <i>Murat Göksedef, Gül Nildem Demir and Şule Gündüz Öğüdücü</i>	50
USING SELF ORGANIZING MAP TO INFER COMMUNITIES IN WEBLOGS' SOCIAL NETWORK <i>Mohsen Jamali, Hassan Abolhassani and Majid Khalilian</i>	58
EXTRACTING KNOWLEDGE TO PREDICT TSP PERFORMANCE ORDER <i>Paula Cecilia Fritzsche, Dolores Rexachs and Emilio Luque</i>	65

MODELING MISSING DATA WITH MARKOV RANDOM FIELDS IN LARGE DATA SETS <i>Esa Junttila and Marko Salmenkivi</i>	73
SELF-ORGANIZING DATA MINING FOR WEATHER FORECASTING <i>Godfrey C. Onwubolu, Petr Buryan, Sitaram Garimella, Visagaperuman Ramachandran, Viti Buadromo and Ajith Abraham</i>	81
INTELLIGENT DATA ANALYSIS USING MULTIPLE CRITERIA DECISION MAKING <i>Crina Grosan and Ajith Abraham</i>	89
MINING RULES IN THE <i>ONTO4AR</i> FRAMEWORK FOR UPDATING DOMAIN ONTOLOGIES <i>Cláudia Antunes</i>	95
MINING FREQUENT INDUCED SUBTREES WITH APPLICATION TO PHYLOGENETIC TREES <i>Marwa M. A. Hussein, Taysir H. A. Soliman and Omar H. Karam</i>	102

SHORT PAPERS

DIMENSIONALITY REDUCTION WITH PARALLEL CORPORA <i>YingJu Xia and HaoYu</i>	113
AN APPROACH FOR THE CONCEPTUAL MODELING OF CLUSTERING MINING IN THE KDD PROCESS <i>Jose Zubcoff and Juan Trujillo</i>	119
FEATURE RANKING BASED ON WEIGHTS ESTIMATED BY MULTIOBJECTIVE OPTIMIZATION <i>Daniela Zaharie, Diana Lungeanu and Stefan Holban</i>	124
ANALYZING EFFICIENT EMBEDDING LIST STRUCTURES IN SUBGRAPH MINING <i>Nivin A. Helal, Taysir H. A. Soliman and Omar H. Karam</i>	129
DETECTING THE COMMUNITY STRUCTURES IN THE GAME OF GO <i>Yuri Malitsky, Christopher Fellows and Gregory Wojtaszczyk</i>	135
REALISTIC SYNTHETIC DATA FOR RULE MINING <i>Colin Cooper and Michele Zito</i>	140
AUTOMATIC TRANSFORMATION OF OWL ONTOLOGIES TO SQL RELATIONAL DATABASES <i>Irina Astrova, Ahto Kalja and Nahum Korda</i>	145
DETECTING INSIDER THREATS WITH MACHINE LEARNING ALGORITHMS <i>Sule Simsek, Ralph W. Wilkerson and R. Joe Stanley</i>	150
EFFICIENT WAY OF SIMULATING PRINCIPAL COMPONENT ANALYSIS FOR SPECIES IDENTIFICATION <i>Fahim Md and Aggarwal Ruchi</i>	155

MF-ARTMAP NEURAL NETWORK PREDICTION OF THE MARKET POSITION OF THE COMPANY <i>Agáta Bodnárová and Kamila Olševičová</i>	160
TEXT MINING SYSTEMS FOR PREDICTING MARKET RESPONSE TO NEWS <i>Marc-André Mittermayer and Gerhard F. Knolmayer</i>	164
SELF-ORGANIZING DATA MINING USING ENHANCED GROUP METHOD DATA HANDLING APPROACH <i>Godfrey C. Onwubolu, Petr Buryan and Ajith Abraham</i>	170
SYSTEM OF ANALYTICAL QUESTIONS and REPORTS ON MINING in health DATA - A CASE STUDY <i>Jan Rauch and Marie Tomečková</i>	176
A KNOWLEDGE MANAGEMENT SYSTEM FOR ORGANIZING MEDLINE DATABASE <i>Su-Shing Chen and Hyunki Kim</i>	182
RANK BASES IN SPACES OF FINITE METRICS <i>Archil Maysuradze</i>	187
EVALUATION OF PROJECTION TECHNIQUES USING HUBERT'S Γ STATISTICS <i>Dorina Marghescu</i>	192
A NEW APPROACH TO IMPROVE MULTI-DIMENSIONAL STOCK DATA REDUCTION <i>Jian Jiang, Zhe Zhang, Huaiqing Wang, Xiaoyan Liu, Xuhao Luo and Lin Wang</i>	198

REFLECTION PAPERS

PROFILING AND ANALYZING E-CHURNERS WITH CLUSTERING <i>Vasilis Aggelis</i>	205
STATE OF THE ART OF OUTLIER DETECTION IN STREAMING DATA <i>Amardeep Kaur, M.P.S. Bhatia and S.M. Bhaskar</i>	209
DATA MINING TECHNIQUES FOR SUSPICIOUS EMAIL DETECTION: A COMPARATIVE STUDY <i>S. Appavu alias Balamurugan, R. Rajaram, G.Athiappan and M. Muthupandian</i>	213

POSTERS

SCADA SYSTEM FOR TEMPERATURE SUPERVISION OF SLAB CASTING MOULD <i>Hilario López, Antonio Robles, Iván Machón, Luis Fernando Sancho and Eva Fernández</i>	221
---	-----

A TEXT MINING SUPPORT TOOL USING THE PROJECT LOOKING GLASS <i>Yuki Makita, Vu Quang and Hitoshi Sasaki</i>	224
GEOSPATIAL DATA MINING PROBLEMS <i>Magdalene Grantson Borgelt</i>	227
A NEW AND OPTIMAL ARCHITECTURE FOR APPLYING ASSOCIATION RULE MINING ALGORITHMS ON GRID <i>Asgarali Bouyer, Bahman Arasteh, Mohammad Bager Karimi, Mohammad Javad Hoseyni, Ali Bouyer and Ali Movaghar</i>	230
AUTHOR INDEX	