

Lecture Notes in Artificial Intelligence 5632

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Petra Perner (Ed.)

Machine Learning and Data Mining in Pattern Recognition

6th International Conference, MLDM 2009
Leipzig, Germany, July 23-25, 2009
Proceedings

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada

Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editor

Petra Perner

Institute of Computer Vision and Applied Computer Sciences (IBal)

Kohlenstr. 2, 04107 Leipzig, Germany

E-mail: pperner@ibai-institut.de

Library of Congress Control Number: 2009930146

CR Subject Classification (1998): I.2, F.4, I.4, I.5, H.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-03069-6 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-03069-7 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12714995 06/3180 5 4 3 2 1 0

Preface

*There is no royal road to science,
and only those who do not dread the fatiguing climb of its steep paths have a chance
of gaining its luminous summits.*

*Karl Marx
A Universal Genius of the 19th Century*

Many scientists from all over the world during the past two years since the MLDM 2007 have come along on the stony way to the sunny summit of science and have worked hard on new ideas and applications in the area of data mining in pattern recognition. Our thanks go to all those who took part in this year's MLDM. We appreciate their submissions and the ideas shared with the Program Committee. We received over 205 submissions from all over the world to the International Conference on Machine Learning and Data Mining, MLDM 2009. The Program Committee carefully selected the best papers for this year's program and gave detailed comments on each submitted paper. There were 63 papers selected for oral presentation and 17 papers for poster presentation.

The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data-mining methods for the different multimedia data types such as image mining, text mining, video mining and Web mining. Among these topics this year were special contributions to subtopics such as attribute discretization and data preparation, novelty and outlier detection, and distances and similarities. The section "Aspects of Data Mining" presented work that considers automating the whole data mining process. The sections also cover data mining in medicine.

Four papers were selected for the Best Paper Award. The final decision on the Best Paper was made during the conference and was based on the review results, the presentation style and the discussion. Please check www.mldm.de for information about the nominated papers and the final decision.

We thank all members of the Institute of Applied Computer Sciences, Leipzig, Germany (www.ibai-institut.de) who handled the conference as secretariat. We appreciate the help and understanding of the editorial staff at Springer, and in particular Alfred Hofmann, who supported the publication of these proceedings in the LNAI series.

Last, but not least, we wish to thank all the speakers and participants who contributed to the success of the conference. We would also like to encourage those who could not get their paper accepted to MLDM 2009 to work out the comments of the reviewers and resubmit their paper to our next conference. See you at MLDM 2011 (www.mldm.de).

International Conference on Machine Learning and Data Mining in Pattern Recognition MLDM 2009

Chair

Petra Perner
Institute of Computer Vision and Applied Computer Sciences IBal Leipzig, Germany

Program Committee

Agnar Aamodt	NTNU, Norway
Jacky Baltes	University of Manitob, Canada
Max Bramer	University of Portsmouth, UK
Horst Bunke	University of Bern, Switzerland
Krzysztof Cios	University of Colorado, USA
Christoph F. Eick	University of Houston, USA
Ana Fred	Technical University of Lisbon, Portugal
Giorgio Giacinto	University of Cagliari, Italy
Makato Haraguchi	Hokkaido University Sapporo, Japan
Robert J. Hilderman	University of Regina, Canada
Tin Kam Ho	Bell Laboratories, USA
Atsushi Imiya	Chiba University, Japan
Horace Ip	City University, Hong Kong
Abraham Kandel	University of South Florida , USA
Dimitrios A. Karras	Chalkis Institute of Technology, Greece
Adam Krzyzak	Concordia University, Montreal, Canada
Longin Jan Latecki	Temple University Philadelphia, USA
Tao Li	Florida International University, USA
Brian Lovell	University of Queensland, Australia
Mariofanna Milanova	University of Arkansas at Little Rock, USA
Thang V. Pham	University of Amsterdam, The Netherlands
Maria da Graca Pimentel	Universidade de São Paulo, Brazil
Petia Radeva	Universitat Autònoma de Barcelona, Spain
Michael Richter	University of Calgary, Canada
Fabio Roli	University of Cagliari, Italy
Linda Shapiro	University of Washington, USA
Sameer Singh	Loughborough University, UK

David Steinberg	Tel Aviv University, Israel
Francesco Tortorella	Universita' degli Studi di Cassino, Italy
Patrick Wang	Northeastern University, USA

Additional Reviewers

Indriyati Atmosukarto	University of Washington, USA
Jiun-Hung Chen	University of Washington, USA
Natalia Larios	University of Washington, USA
Dingding Liu	University of Washington, USA
Marta Penas Centeno	University of Washington, USA
Jia Wu	University of Washington, USA
Sara Rolfe	University of Washington, USA
Shulin Yang	University of Washington, USA
Mabel Raza	University of Washington, USA
Rosalia Tungaraza	University of Washington, USA
Katarzyna Wilamowska	University of Washington, USA
Tore Amble	NTNU, Norway
Richard Blake	NTNU, Norway
Tore Bruland	NTNU, Norway
Tor Gunnar Houeland	NTNU, Norway
Dung Manh Chu	University of Amsterdam, The Netherlands
Francesco Fontanella	Universita' degli Studi di Cassino, Italy
Claudio Marrocco	Universita' degli Studi di Cassino, Italy
Mario Molinara	Universita' degli Studi di Cassino, Italy
Maria Teresa Ricamato	Universita' degli Studi di Cassino, Italy
Alessandra Scotto di Freca	Universita' degli Studi di Cassino, Italy
Paolo Simeone	Universita' degli Studi di Cassino, Italy
Davide Ariu	University of Cagliari, Italy
Battista Biggio	University of Cagliari, Italy
Luca Didaci	University of Cagliari, Italy
Giorgio Fumera	University of Cagliari, Italy
Gian Luca Marcialis	University of Cagliari, Italy
Daniele Muntoni	University of Cagliari, Italy
Ignazio Pillai	University of Cagliari, Italy
Luca Piras	University of Cagliari, Italy
Ajita Rattani	University of Cagliari, Italy
Roberto Tronci	University of Cagliari, Italy

Table of Contents

Attribute Discretization and Data Preparation

Improved Comprehensibility and Reliability of Explanations via Restricted Halfspace Discretization	1
<i>Klaus Truemper</i>	
Selection of Subsets of Ordered Features in Machine Learning	16
<i>O. Seredin, A. Kopylov, and V. Mottl</i>	
Combination of Vector Quantization and Visualization	29
<i>Olga Kurasova and Alma Molyté</i>	
Discretization of Target Attributes for Subgroup Discovery	44
<i>Katherine Moreland and Klaus Truemper</i>	
Preserving Privacy in Time Series Data Classification by Discretization	53
<i>Ye Zhu, Yongjian Fu, and Huirong Fu</i>	
Using Resampling Techniques for Better Quality Discretization	68
<i>Taimur Qureshi and Djamel A. Zighed</i>	

Classification

A Large Margin Classifier with Additional Features	82
<i>Xinwang Liu, Jianping Yin, En Zhu, Guomin Zhang, Yubin Zhan, and Miaomiao Li</i>	
Sequential EM for Unsupervised Adaptive Gaussian Mixture Model Based Classifier	96
<i>Bashar Awwad Shiekh Hasan and John Q. Gan</i>	
Optimal Double-Kernel Combination for Classification	107
<i>Feng Wang and Hongbin Zhang</i>	
Efficient AdaBoost Region Classification	123
<i>M. Moed and E.N. Smirnov</i>	
A Linear Classification Method in a Very High Dimensional Space Using Distributed Representation	137
<i>Takao Kobayashi and Ikuko Shimizu</i>	
PMCRI: A Parallel Modular Classification Rule Induction Framework	148
<i>Frederic Stahl, Max Bramer, and Mo Adda</i>	

Dynamic Score Combination: A Supervised and Unsupervised Score
Combination Method 163
Roberto Tronci, Giorgio Giacinto, and Fabio Roli

ODDboost: Incorporating Posterior Estimates into AdaBoost 178
Olga Barinova and Dmitry Vetrov

Ensemble Classifier Learning

Ensemble Learning: A Study on Different Variants of the Dynamic
Selection Approach 191
*João Mendes-Moreira, Alipio Mario Jorge, Carlos Soares, and
Jorge Freire de Sousa*

Relevance and Redundancy Analysis for Ensemble Classifiers 206
Rakkrit Duangsoithong and Terry Windeatt

Drift-Aware Ensemble Regression 221
*Frank Rosenthal, Peter Benjamin Volk, Martin Hahmann,
Dirk Habich, and Wolfgang Lehner*

Concept Drifting Detection on Noisy Streaming Data in Random
Ensemble Decision Trees 236
Peipei Li, Xuegang Hu, Qianhui Liang, and Yunjun Gao

Association Rules and Pattern Mining

Mining Multiple Level Non-redundant Association Rules through
Two-Fold Pruning of Redundancies 251
Corrado Loglisci and Donato Malerba

Pattern Mining with Natural Language Processing: An Exploratory
Approach 266
Ana Cristina Mendes and Cláudia Antunes

Is the Distance Compression Effect Overstated? Some Theory and
Experimentation 280
Stephen France and Douglas Carroll

Support Vector Machines

Fast Local Support Vector Machines for Large Datasets 295
Nicola Segata and Enrico Blanzieri

The Effect of Domain Knowledge on Rule Extraction from Support
Vector Machines 311
Nahla Barakat and Andrew P. Bradley

Towards B-Coloring of SOM.....	322
<i>Haytham Elghazel and Khalid Benabdeslem</i>	

Clustering

CSBIterKmeans: A New Clustering Algorithm Based on Quantitative Assessment of the Clustering Quality	337
<i>Tarek Smaoui, Sascha Müller, and Christian Müller-Schloer</i>	
Agent-Based Non-distributed and Distributed Clustering	347
<i>Ireneusz Czarnowski and Piotr Jędrzejowicz</i>	
An Evidence Accumulation Approach to Constrained Clustering Combination	361
<i>Daniel Duarte Abdala and Xiaoyi Jiang</i>	
Fast Spectral Clustering with Random Projection and Sampling	372
<i>Tomoya Sakai and Atsushi Imiya</i>	
How Much True Structure Has Been Discovered?: Validating Explorative Clustering on a Hold-Out Test Set	385
<i>F. Höppner</i>	
Efficient Clustering of Web-Derived Data Sets	398
<i>Luís Sarmiento, Alexander Kehlenbeck, Eugénio Oliveira, and Lyle Ungar</i>	
A Probabilistic Approach for Constrained Clustering with Topological Map.....	413
<i>Khalid Benabdeslem and Jihene Snoussi</i>	

Novelty and Outlier Detection

Relational Frequent Patterns Mining for Novelty Detection from Data Streams	427
<i>Michelangelo Ceci, Annalisa Appice, Corrado Loglisci, Costantina Caruso, Fabio Fumarola, Carmine Valente, and Donato Malerba</i>	
A Comparative Study of Outlier Detection Algorithms.....	440
<i>Charlie Isaksson and Margaret H. Dunham</i>	
Outlier Detection with Explanation Facility	454
<i>Manuel Mejía-Lavalle and Atlántida Sánchez Vivar</i>	

Learning

Concept Learning from (Very) Ambiguous Examples 465
Dominique Bouthinon, Henry Soldano, and Véronique Ventos

Finding Top-*N* Pseudo Formal Concepts with Core Intents 479
Yoshiaki Okubo and Makoto Haraguchi

On Fixed Convex Combinations of No-Regret Learners 494
Jan-P. Calliess

An Improved Tabu Search (ITS) Algorithm Based on Open Cover
 Theory for Global Extremums 505
Kemal Yüeksek and Serhat Cakaloglu

The Needles-in-Haystack Problem 516
Katherine Moreland and Klaus Truemper

Data Mining on Multimedia Data

An Evidence-Driven Probabilistic Inference Framework for Semantic
 Image Understanding 525
*Spiros Nikolopoulos, Georgios Th. Papadopoulos,
 Ioannis Kompatsiaris, and Ioannis Patras*

Detection of Masses in Mammographic Images Using Simpson’s
 Diversity Index in Circular Regions and SVM 540
*André Pereira Nunes, Aristófaes Corrêa Silva, and
 Anselmo Cardoso de Paiva*

Mining Lung Shape from X-Ray Images 554
Vassili Kovalev, Aliaksandr Prus, and Pavel Vankevich

A Wavelet-Based Method for Detecting Seismic Anomalies in Remote
 Sensing Satellite Data 569
Pan Xiong, Yaxin Bi, and Xuhui Shen

Spectrum Steganalysis of WAV Audio Streams 582
Qingzhong Liu, Andrew H. Sung, and Mengyu Qiao

Audio-Based Emotion Recognition in Judicial Domain: A Multilayer
 Support Vector Machines Approach 594
E. Fersini, E. Messina, G. Arosio, and F. Archetti

Learning with a Quadruped Chopstick Robot 603
Wei-Chung Lee, Jong-Chen Chen, Shou-zhe Wu, and Kuo-Ming Lin

Dissimilarity Based Vector Space Embedding of Graphs Using
 Prototype Reduction Schemes 617
Kaspar Riesen and Horst Bunke

Text Mining

Using Graph-Kernels to Represent Semantic Information in Text Classification	632
<i>Teresa Gonçalves and Paulo Quaresma</i>	
A General Framework of Feature Selection for Text Categorization	647
<i>Hongfang Jing, Bin Wang, Yahui Yang, and Yan Xu</i>	
New Semantic Similarity Based Model for Text Clustering Using Extended Gloss Overlaps	663
<i>Walaa K. Gad and Mohamed S. Kamel</i>	

Aspects of Data Mining

Learning Betting Tips from Users' Bet Selections	678
<i>Erik Štrumbelj, Marko Robnik Šikonja, and Igor Kononenko</i>	
An Approach to Web-Scale Named-Entity Disambiguation	689
<i>Luís Sarmiento, Alexander Kehlenbeck, Eugénio Oliveira, and Lyle Ungar</i>	
A General Learning Method for Automatic Title Extraction from HTML Pages	704
<i>Sahar Changuel, Nicolas Labroche, and Bernadette Bouchon-Meunier</i>	
Regional Pattern Discovery in Geo-referenced Datasets Using PCA	719
<i>Oner Ulvi Celepcikay, Christoph F. Eick, and Carlos Ordonez</i>	
Memory-Based Modeling of Seasonality for Prediction of Climatic Time Series	734
<i>Daniel Nikovski and Ganesan Ramachandran</i>	
A Neural Approach for SME's Credit Risk Analysis in Turkey	749
<i>Gölnur Derelioğlu, Fikret Gürgeç, and Nesrin Okay</i>	
Assisting Data Mining through Automated Planning	760
<i>Fernando Fernández, Daniel Borrajo, Susana Fernández, and David Manzano</i>	
Predictions with Confidence in Applications	775
<i>Mikhail Dashevskiy and Zhiyuan Luo</i>	

Data Mining in Medicine

Aligning Bayesian Network Classifiers with Medical Contexts	787
<i>Linda C. van der Gaag, Silja Renooij, Ad Feelders, Arend de Groote, Marinus J.C. Eijkemans, Frank J. Broekmans, and Bart C.J.M. Fauser</i>	

Assessing the Eligibility of Kidney Transplant Donors	802
<i>Francisco Reinaldo, Carlos Fernandes, Md. Anishur Rahman, Andreia Malucelli, and Rui Camacho</i>	
Lung Nodules Classification in CT Images Using Simpson's Index, Geometrical Measures and One-Class SVM	810
<i>Cleriston Araujo da Silva, Aristófanes Corrêa Silva, Stelmo Magalhães Barros Netto, Anselmo Cardoso de Paiva, Geraldo Braz Junior, and Rodolfo Acatauassú Nunes</i>	
Author Index	823