

Lecture Notes in Artificial Intelligence 6871

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann

DFKI and Saarland University, Saarbrücken, Germany

Petra Perner (Ed.)

Machine Learning and Data Mining in Pattern Recognition

7th International Conference, MLDM 2011

New York, NY, USA

August 30 – September 3, 2011

Proceedings



Springer

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, IBaI

Kohlenstraße 2, 04107 Leipzig, Germany

E-mail: pperner@ibai-institut.de

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-23198-8

e-ISBN 978-3-642-23199-5

DOI 10.1007/978-3-642-23199-5

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011933947

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

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2011

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The seventh event of the International Conference on Machine Learning and Data Mining (MLDM) was held in New York (www.mldm.de) running under the umbrella of the world congress "The Frontiers in Intelligent Data and Signal Analysis, DSA2011."

For this edition the Program Committee received 170 submissions. After the peer-review process, we accepted 56 high-quality papers for oral presentation, and from these 44 are included in this proceedings book. 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. Extended versions of selected papers will appear in the international journal Transactions on Machine Learning and Data Mining (www.ibai-publishing.org/journal/mldm).

Fourteen papers were selected for poster presentation and are published in the *MLDM Poster Proceedings by ibai-publishing* (www.ibai-publishing.org).

A tutorial on Data Mining and a tutorial on Case-Based Reasoning were held before the conference.

We were pleased to give out the best paper award for MLDM for the third time this year. The final decision was made by the Best Paper Award Committee based on the presentation by the authors and the discussion with the auditorium. The ceremony took place at the end of the conference. This prize is sponsored by ibai solutions (www.ibai-solutions.de), one of the leading companies in data mining for marketing, Web mining and e-commerce.

The conference was rounded up by an outlook of new challenging topics in machine learning and data mining before the Best Paper Award Ceremony.

We thank the 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. St. Petersburg will host the next world congress on "The Frontiers in Intelligent Data and Signal Analysis, DSA2012" (www.worldcongressdsa.com) in 2012, combining under its roof the following events: International Conference on Machine Learning and Data Mining (MLDM); the Industrial Conference on Data Mining (ICDM), and the International Conference on Mass Data Analysis of Signals and Images in Medicine, Biotechnology, Chemistry and Food Industry (MDA).

Organization

Chair

Petra Perner

IBaI Leipzig, Germany

Program Committee

Agnar Aamodt	NTNU, Norway
Jacky Baltes	University of Manitoba, Canada
Max Bramer	University of Portsmouth, UK
Horst Bunke	University of Bern, Switzerland
Krzysztof Cios	University of Colorado, USA
John Debenham	University of Technology, Australia
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
Lukasz Kurgan	University of Alberta, 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 Sao 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	Università degli Studi di Cassino, Italy
Patrick Wang	Northeastern University, USA

Additional Reviewers

Richard Blake	NTNU, Norway
Gleb Sizov	NTNU, Norway
Pinar Özturk	NTNU, Norway
Pål Sætrom (Paal Saetrom)	NTNU, Norway
Helge Langseth	NTNU, Norway
Tore Bruland	NTNU, Norway
Ajita Rattani	University of Cagliari, Italy
Ignazio Pillai	University of Cagliari, Italy
Davide Ariu	University of Cagliari, Italy
Luca Didaci	University of Cagliari, Italy
Roberto Tronci	University of Cagliari, Italy
Battista Biggio	University of Cagliari, Italy

Table of Contents

Classification and Decision Theory

Quadratically Constrained Maximum a Posteriori Estimation for Binary Classifier	1
<i>Tatsuya Yokota and Yukihiro Yamashita</i>	
Hubness-Based Fuzzy Measures for High-Dimensional k -Nearest Neighbor Classification	16
<i>Nenad Tomašev, Miloš Radovanović, Dunja Mladenić, and Mirjana Ivanović</i>	
Decisions: Algebra and Implementation	31
<i>Antonina Danylenko, Jonas Lundberg, and Welf Löwe</i>	
Smoothing Multinomial Naïve Bayes in the Presence of Imbalance	46
<i>Alexander Y. Liu and Cheryl E. Martin</i>	
ACE-Cost: Acquisition Cost Efficient Classifier by Hybrid Decision Tree with Local SVM Leaves	60
<i>Liyun Li, Umut Topkara, and Nasir Memon</i>	
Informative Variables Selection for Multi-relational Supervised Learning	75
<i>Dhafer Lahbib, Marc Boullé, and Dominique Laurent</i>	
Separability of Split Value Criterion with Weighted Separation Gains ...	88
<i>Krzysztof Grąbczewski</i>	
Granular Instances Selection for Fuzzy Modeling	99
<i>S. Sakinah S. Ahmad and Witold Pedrycz</i>	
Parameter-Free Anomaly Detection for Categorical Data	112
<i>Shu Wu and Shengrui Wang</i>	
Fuzzy Semi-supervised Support Vector Machines	127
<i>Houda Benbrahim</i>	
GENCCS: A Correlated Group Difference Approach to Contrast Set Mining	140
<i>Mondelle Simeon and Robert Hilderman</i>	
Collective Classification Using Heterogeneous Classifiers	155
<i>Zehra Cataltepe, Abdullah Sonmez, Kadriye Baglioglu, and Ayse Erzan</i>	

Spherical Nearest Neighbor Classification: Application to Hyperspectral Data	170
<i>Dalton Lunga and Okan Ersoy</i>	
Adaptive Kernel Diverse Density Estimate for Multiple Instance Learning	185
<i>Tao Xu, Iker Gondra, and David Chiu</i>	
Boosting Inspired Process for Improving AUC	199
<i>Victor S. Sheng and Rahul Tada</i>	

Theory of Learning

Investigation in Transfer Learning: Better Way to Apply Transfer Learning between Agents	210
<i>Luiz Antonio Celiberto Junior and Jackson P. Matsuura</i>	
Exploration Strategies for Learned Probabilities in Smart Terrain	224
<i>John Sullins</i>	
Sensitivity Analysis for Weak Constraint Generation	239
<i>Jamshaid G. Mohebzada, Michael M. Richter, and Guenther Ruhe</i>	
Dictionary Learning Based on Laplacian Score in Sparse Coding	253
<i>Jin Xu and Hong Man</i>	

Clustering

A Practical Approach for Clustering Transaction Data	265
<i>Mohamed Bouguessa</i>	
Hierarchical Clustering with High Order Dissimilarities	280
<i>Helena Aidos and Ana Fred</i>	
Clust-XPaths: Clustering of XML Paths	294
<i>Amina Madani, Omar Boussaid, and Djamel Eddine Zegour</i>	
Comparing Clustering and Metaclustering Algorithms	306
<i>Elio Lozano and Edgar Acuña</i>	

Applications in Medicine

Detection of Phenotypes in Microarray Data Using Force-Directed Placement Transforms	320
<i>Dragana Veljkovic Perez and Kay A. Robbins</i>	
On the Temporal Behavior of EEG Recorded during Real Finger Movement	335
<i>Bashar Awwad Shiekh Hasan</i>	

A Machine Learning and Data Mining Framework to Enable Evolutionary Improvement in Trauma Triage	348
<i>Douglas A. Talbert, Matt Honeycutt, and Steve Talbert</i>	
A Decision Support System Based on the Semantic Analysis of Melanoma Images Using Multi-elitist PSO and SVM	362
<i>Weronika Pigtkowska, Jerzy Martyna, Leszek Nowak, and Karol Przystalski</i>	

WebMining/Information Mining

Authorship Similarity Detection from Email Messages	375
<i>Xiaoling Chen, Peng Hao, R. Chandramouli, and K.P. Subbalakshmi</i>	
An Investigation Concerning the Generation of Text Summarisation Classifiers Using Secondary Data	387
<i>Matias Garcia-Constantino, Frans Coenen, P.-J. Noble, Alan Radford, Christian Setzkorn, and Aine Tierney</i>	
Comparing the One-Vs-One and One-Vs-All Methods in Benthic Macroinvertebrate Image Classification	399
<i>Henry Joutsijoki and Martti Juhola</i>	
Incremental Web-Site Boundary Detection Using Random Walks	414
<i>Ayesh Alshukri, Frans Coenen, and Michele Zito</i>	
Discovering Text Patterns by a New Graphic Model	428
<i>Minhua Huang and Robert M. Haralick</i>	
Topic Sentiment Change Analysis	443
<i>Yu Jiang, Weiyi Meng, and Clement Yu</i>	
Adaptive Context Modeling for Deception Detection in Emails	458
<i>Peng Hao, Xiaoling Chen, Na Cheng, R. Chandramouli, and K.P. Subbalakshmi</i>	
Contrasting Correlations by an Efficient Double-Clique Condition	469
<i>Aixiang Li, Makoto Haraguchi, and Yoshiaki Okubo</i>	

Machine Learning and Image Mining

Estimating Image Segmentation Difficulty	484
<i>Dingding Liu, Yingen Xiong, Kari Pulli, and Linda Shapiro</i>	
Mining Spatial Trajectories Using Non-parametric Density Functions ...	496
<i>Chun-Sheng Chen, Christoph F. Eick, and Nouhad J. Rizk</i>	

Exploring Synergetic Effects of Dimensionality Reduction and
Resampling Tools on Hyperspectral Imagery Data Classification 511
J.S. Sánchez, V. García, and R.A. Mollineda

A Comparison between Haralick's Texture Descriptor and the Texture
Descriptor Based on Random Sets for Biological Images..... 524
Anja Attig and Petra Perner

Time Series and Frequent Item Set Mining

Unsupervised Discovery of Motifs under Amplitude Scaling and Shifting
in Time Series Databases 539
Tom Armstrong and Eric Drewniak

Static Load Balancing of Parallel Mining of Frequent Itemsets Using
Reservoir Sampling..... 553
Robert Kessl

GA-TVRC: A Novel Relational Time Varying Classifier to Extract
Temporal Information Using Genetic Algorithms 568
İsmail Güneş, Zehra Çataltepe, and Şule Gündüz Öğüdücü

Aspects of Machine Learning and Data Mining

Detection of Communities and Bridges in Weighted Networks..... 584
Tanwistha Saha, Carlotta Domeniconi, and Huzeefa Rangwala

Techniques for Improving Filters in Power Grid Contingency
Analysis 599
*Robert Adolf, David Haglin, Mahantesh Halappanavar,
Yousu Chen, and Zhenyu Huang*

Author Index 613