Lecture Notes in Artificial Intelligence 4065

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Petra Perner (Ed.)

Advances in Data Mining

Applications in Medicine, Web Mining, Marketing, Image and Signal Mining

6th Industrial Conference on Data Mining, ICDM 2006
Leipzig, Germany, July 14-15, 2006
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editor

Petra Perner
Institute of Computer Vision and Applied Computer Sciences, IBaI
Körnerstr. 10, 04107 Leipzig, Germany
E-mail: pperner@ibai-institut.de

Library of Congress Control Number: 2006928502

CR Subject Classification (1998): I.2.6, I.2, H.2.8, K.4.4, J.3, I.4, J.6, J.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-36036-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-36036-0 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 is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11790853 06/3142 5 4 3 2 1 0

Preface

The Industrial Conference on Data Mining ICDM-Leipzig was the sixth event in a series of annual events which started in 2000. We are pleased to note that the topic data mining with special emphasis on real-world applications has been adopted by so many researchers all over the world into their research work. We received 156 papers from 19 different countries.

The main topics are data mining in medicine and marketing, web mining, mining of images and signals, theoretical aspects of data mining, and aspects of data mining that bundle a series of different data mining applications such as intrusion detection, knowledge management, manufacturing process control, time-series mining and criminal investigations.

The Program Committee worked hard in order to select the best papers. The acceptance rate was 30%. All these selected papers are published in this proceedings volume as long papers up to 15 pages. Moreover we installed a forum where work in progress was presented. These papers are collected in a special poster proceedings volume and show once more the potentials and interesting developments of data mining for different applications.

Three new workshops have been established in connection with ICDM: (1) Mass Data Analysis on Images and Signals, MDA 2006; (2) Data Mining for Life Sciences, DMLS 2006; and (3) Data Mining in Marketing, DMM 2006. These workshops are developing new topics for data mining under the aspect of the special application. We are pleased to see how many interesting developments are going on in these fields.

We would like to express our appreciation to the reviewers for their precise and highly professional work. 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.

We wish to thank all speakers, participants, and industrial exhibitors who contributed to the success of the conference.

We are looking forward to welcoming you to ICDM 2007 (www.data-mining-forum.de) and to the new work presented there.

July 2006 Petra Perner

Table of Contents

Data Mining in Medicine

Using Prototypes and Adaptation Rules for Diagnosis of Dysmorphic Syndromes
Rainer Schmidt, Tina Waligora 1

OVA Scheme vs. Single Machine Approach in Feature Selection for Microarray Datasets
Chia Huey Ooi, Madhu Chetty, Shyh Wei Teng 10

Similarity Searching in DNA Sequences by Spectral Distortion Measures
Tuan Duc Pham 24

Multispecies Gene Entropy Estimation, a Data Mining Approach
Xiaoxu Han 38

A Unified Approach for Discovery of Interesting Association Rules in Medical Databases
Harleen Kaur, Siri Krishan Wasan, Ahmed Sultan Al-Hegami, Vasudha Bhatnagar 53

Named Relationship Mining from Medical Literature
Isabelle Bichindaritz 64

Experimental Study of Evolutionary Based Method of Rule Extraction from Neural Networks in Medical Data
Urszula Markowska-Kaczmar, Rafal Matkowski 76

Web Mining and Logfile Analysis

HTTPHunting: An IBR Approach to Filtering Dangerous HTTP Traffic
Florentino Fdez-Riverola, Lourdes Borrajo, Rosalia Laza, Francisco J. Rodríguez, David Martínez 91

A Comparative Performance Study of Feature Selection Methods for the Anti-spam Filtering Domain
Jose Ramon Méndez, Florentino Fdez-Riverola, Fernando Díaz, Eva Lorenzo Iglesias, Juan Manuel Corchado 106

Evaluation of Web Robot Discovery Techniques: A Benchmarking Study
Nick Geens, Johan Huysmans, Jan Vanthienen 121

Data Preparation of Web Log Files for Marketing Aspects Analyses
Meike Reichle, Petra Perner, Klaus-Dieter Althoff 131

UP-DRES: User Profiling for a Dynamic REcommendation System
Enza Messina, Daniele Toscani, Francesco Archetti 146

Improving Effectiveness on Clickstream Data Mining
Cristina Wanzeller, Orlando Belo 161

Conceptual Knowledge Retrieval with FooCA: Improving Web Search Engine Results with Contexts and Concept Hierarchies
Bjoern Koester 176

Theoretical Aspects of Data Mining

A Pruning Based Incremental Construction Algorithm of Concept Lattice
Ji-Fu Zhang, Li-Hua Hu, Su-Lan Zhang 191

Association Rule Mining with Chi-Squared Test Using Alternate Genetic Network Programming
Kaoru Shimada, Kotaro Hirasawa, Jinglu Hu 202

Ordinal Classification with Monotonicity Constraints
Tomáš Horváth, Peter Vojtáš 217

Local Modelling in Classification on Different Feature Subspaces
Gero Szepannek, Claus Weihs 226

Supervised Selection of Dynamic Features, with an Application to Telecommunication Data Preparation
Sylvain Ferrandiz, Marc Boullé 239

Using Multi-SOMs and Multi-Neural-Gas as Neural Classifiers
Nils Goerke, Alexandra Scherbart 250

Derivative Free Stochastic Discrete Gradient Method with Adaptive Mutation
Ranadhir Ghosh, Moumita Ghosh, Adil Bagirov 264

Data Mining in Marketing

Association Analysis of Customer Services from the Enterprise Customer Management System
Sung-Ju Kim, Dong-Sik Yun, Byung-Soo Chang 279

Feature Selection in an Electric Billing Database Considering Attribute Inter-dependencies
Manuel Mejía-Lavalle, Eduardo F. Morales 284

Learning the Reasons Why Groups of Consumers Prefer Some Food Products
Juan José del Coz, Jorge Díez, Antonio Bahamonde, Carlos Sañudo, Matilde Alfonso, Philippe Berge, Eric Dransfield, Costas Stamataris, Demetrios Zygoyiannis, Tyri Valdimarsdottir, Edi Piasentier, Geoffrey Nute, Alan Fisher 297

Exploiting Randomness for Feature Selection in Multinomial Logit: A CRM Cross-Sell Application
Anita Prinzie, Dirk Van den Poel 310

Data Mining Analysis on Italian Family Preferences and Expenditures
Paola Annoni, Pier Alda Ferrari, Silvia Salini 324

Multiobjective Evolutionary Induction of Subgroup Discovery Fuzzy Rules: A Case Study in Marketing
Francisco Berlanga, María José del Jesus, Pedro González, Francisco Herrera, Mikel Mesonero 337

A Scatter Search Algorithm for the Automatic Clustering Problem
Rasha Shaker Abdule-Wahab, Nicolas Monmarché, Mohamed Slimane, Moaid A. Fahdil, Hilal H. Saleh 350

Multi-objective Parameters Selection for SVM Classification Using NSGA-II
Li Xu, Chunping Li 365

Effectiveness Evaluation of Data Mining Based IDS
Agustín Orfila, Javier Carbó, Arturo Ribagorda 377

Mining Signals and Images

Spectral Discrimination of Southern Victorian Salt Tolerant Vegetation
Chris Matthews, Rob Clark, Leigh Callinan 389

A Generative Graphical Model for Collaborative Filtering of Visual Content
Sabri Boutemedjet, Djemel Ziou 404

A Variable Initialization Approach to the EM Algorithm for Better Estimation of the Parameters of Hidden Markov Model Based Acoustic Modeling of Speech Signals
Md. Shamsul Huda, Ranadhir Ghosh, John Yearwood 416

Mining Dichromatic Colours from Video
Vassili A. Kovalev 431

Feature Analysis and Classification of Classical Musical Instruments: An Empirical Study
Christian Simmermacher, Da Deng, Stephen Cranefield 444

Automated Classification of Images from Crystallisation Experiments
Julie Wilson 459

Aspects of Data Mining

An Efficient Algorithm for Frequent Itemset Mining on Data Streams
Zhi-jun Xie, Hong Chen, Cuiping Li 474

Discovering Key Sequences in Time Series Data for Pattern Classification
Peter Funk, Ning Xiong 492

Data Alignment Via Dynamic Time Warping as a Prerequisite for Batch-End Quality Prediction
Geert Gins, Jairo Espinosa, Ilse Y. Smets, Wim Van Brempt, Jan F.M. Van Impe 506

A Distance Measure for Determining Similarity Between Criminal Investigations
Tim K. Cocx, Walter A. Kosters 511

Establishing Fraud Detection Patterns Based on Signatures
Pedro Ferreira, Ronnie Alves, Orlando Belo, Luís Cortesão 526

Intelligent Information Systems for Knowledge Work(ers)
Klaus-Dieter Althoff, Björn Decker, Alexandre Hanft, Jens Mänz, Régis Newo, Markus Nick, Jörg Rech, Martin Schaaf 539

Nonparametric Approaches for e-Learning Data
Paolo Baldini, Silvia Figini, Paolo Giudici 548

An Intelligent Manufacturing Process Diagnosis System Using Hybrid Data Mining
Joon Hur, Hongchul Lee, Jun-Geol Baek 561

Computer Network Monitoring and Abnormal Event Detection Using Graph Matching and Multidimensional Scaling
Horst Bunke, Peter Dickinson, Andreas Humm, Christophe Irniger, Miro Kraetzl 576

Author Index 591