

Lecture Notes in Artificial Intelligence 4597

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Petra Perner (Ed.)

Advances in Data Mining

Theoretical Aspects and Applications

7th Industrial Conference, ICDM 2007
Leipzig, Germany, July 14-18, 2007
Proceedings

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)
Arno-Nitzsche-Str. 43, 04277 Leipzig, Germany
E-mail: pperner@ibai-institut.de

Library of Congress Control Number: 2007929837

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 3-540-73434-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73434-5 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 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12086436 06/3180 5 4 3 2 1 0

Preface



ICDM / MLDM Medaillie (limited edition)
Meissner Porcellan, the “White Gold” of King
August the Strongest of Saxonia

ICDM 2007 was the seventh event in the Industrial Conference on Data Mining series and was held in Leipzig (www.data-mining-forum.de).

For this edition the Program Committee received 96 submissions from 24 countries (see Fig. 1).

After the peer-review process, we accepted 25 high-quality papers for oral presentation that are included in this proceedings book. The topics range from aspects of classification and prediction, clustering, Web mining, data mining in medicine, applications of data mining, time series and frequent pattern mining, and association rule mining.

Germany	9,30%	4,17%	China	9,30%	1,04%	South Korea	6,98%	3,13%
Czech Republic	6,98%	3,13%	USA	6,98%	2,08%	UK	4,65%	2,08%
Portugal	4,65%	2,08%	Iran	4,65%	2,08%	India	4,65%	2,08%
Brazil	4,65%	1,04%	Hungary	4,65%	1,04%	Mexico	4,65%	1,04%
Finland	2,33%	1,04%	Ireland	2,33%	1,04%	Slovenia	2,33%	1,04%
France	2,33%	1,04%	Israel	2,33%	1,04%	Spain	2,33%	1,04%
Greece	2,33%	1,04%	Italy	2,33%	1,04%	Sweden	2,33%	1,04%
Netherlands	2,33%	1,04%	Malaysia	2,33%	1,04%	Turkey	2,33%	1,04%

Fig. 1. Distribution of papers among countries

Twelve papers were selected for poster presentations that are published in the ICDM Poster Proceedings Volume.

In conjunction with ICDM two workshops were run on special hot application-oriented topics in data mining. The workshop Data Mining in Life Science DMLS 2007 was held the second time this year and the workshop Data Mining in Marketing

DMM 2007 was held for first time this year. Right after ICDM, the International Conference on Machine Learning and Data Mining, MLDM 2007, was held in Leipzig (www.mldm.de)

The invited talk was given by Prof. Richter, titled “Case-Based Reasoning and the Search for Knowledge.” The talk illustrated that case-based reasoning on the lower, i.e., more personal levels is quite useful, in particular in comparison with traditional information-retrieval methods.

We saw an increasing number of industrial participants at our conference in the special sessions that covered topics that are important for industry. An invited talk was given by Andrea Ahlemeyer on the topic of “How to Combine Data Mining and Market-Research Technologies?.” A discussion forum that described and discussed the occupational image of the data miner was given by Prof. Gentsch from the Business Intelligence Group Inc. Special talks were given by industry staff in a marketing workshop that described the special problems of different industries.

We are pleased to announce that we gave out the best paper award for ICDM for a second time this year.

We also established an MLDM/ICDM/MDA Conference Summary Volume first the time this year, which summarizes the vision of the three conferences and the paper presentations and also provides a “Who is Who” in machine learning and data mining by giving each author the chance to present himself.

We thank 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.

July 2007

Petra Perner

Industrial Conference on Data Mining, ICDM 2007

Chair

Petra Perner

IBaI Leipzig, Germany

Committee

Klaus-Peter Adlassnig

Andrea Ahlemeyer-Stubbe

Klaus-Dieter Althoff

Chid Apte

Isabelle Bichindaritz

Leon Bobrowski

Marc Boullé

Juan M. Corchado

Da Deng

Peter Funk

Ron Kenett

Eduardo F. Morales

Stefania Montani

Eric Pauwels

Rainer Schmidt

Stijn Viaene

Rob A. Vingerhoeds

Medical University of Vienna, Austria

ECDM, Gengenbach, Germany

University of Hildesheim, Germany

IBM Yorktown Heights, USA

University of Washington, USA

Bialystok Technical University, Poland

France Télécom, France

Universidad de Salamanca, Spain

University of Otago, New Zealand

Mälardalen University, Sweden

KPA Ltd., Israel

INAOE, Ciencias Computacionales, Mexico

Università del Piemonte Orientale, Italy

CWI Utrecht, The Netherlands

University of Rostock, Germany

KU Leuven, Belgium

Ecole Nationale d'Ingénieurs de Tarbes, France

Additional Reviewers

Fabrice Clerot

Francoise Fessant

Carine Hue

Vincent Lemaire

France Télécom R&D

France Télécom R&D

France Télécom R&D

France Télécom R&D

Table of Contents

Invited Talk

Case Based Reasoning and the Search for Knowledge	1
<i>Michael M. Richter</i>	

Aspects of Classification and Prediction

Subsets More Representative Than Random Ones	15
<i>Ilia Nouretdinov</i>	
Concepts for Novelty Detection and Handling Based on a Case-Based Reasoning Process Scheme	21
<i>Petra Perner</i>	
An Efficient Algorithm for Instance-Based Learning on Data Streams . . .	34
<i>Jürgen Beringer and Eyke Hüllermeier</i>	
Softening the Margin in Discrete SVM	49
<i>Carlotta Orsenigo and Carlo Vercellis</i>	
Feature Selection Using Ant Colony Optimization (ACO): A New Method and Comparative Study in the Application of Face Recognition System	63
<i>Hamidreza Rashidy Kanan, Karim Faez, and Sayyed Mostafa Taheri</i>	
Outlier Detection with Streaming Dyadic Decomposition	77
<i>Chetan Gupta and Robert Grossman</i>	
VISRED –Numerical Data Mining with Linear and Nonlinear Techniques	92
<i>Antonio Dourado, Edgar Ferreira, and Paulo Barbeiro</i>	

Clustering

Clustering by Random Projections	107
<i>Thierry Urruty, Chabane Djeraba, and Dan A. Simovici</i>	
Lightweight Clustering Technique for Distributed Data Mining Applications	120
<i>Lamine M. Aouad, Nhien-An Le-Khac, and Tahar M. Kechadi</i>	

Web Mining

Predicting Page Occurrence in a Click-Stream Data: Statistical and Rule-Based Approach	135
<i>Petr Berka and Martin Labský</i>	
Improved IR in Cohesion Model for Link Detection System	148
<i>Krishnamurthy Lakshmi and Saswati Mukherjee</i>	
Improving a State-of-the-Art Named Entity Recognition System Using the World Wide Web	163
<i>Richárd Farkas, György Szarvas, and Róbert Ormándi</i>	

Data Mining in Medicine

ISOR-2: A Case-Based Reasoning System to Explain Exceptional Dialysis Patients	173
<i>Olga Vorobieva, Alexander Rumyantsev, and Rainer Schmidt</i>	
The Role of Prototypical Cases in Biomedical Case-Based Reasoning ...	184
<i>Isabelle Bichindaritz</i>	

Applications of Data Mining

A Search Space Reduction Methodology for Large Databases: A Case Study	199
<i>Angel Kuri-Morales and Fátima Rodríguez-Erazo</i>	
Combining Traditional and Neural-Based Techniques for Ink Feed Control in a Newspaper Printing Press	214
<i>Cristofer Englund and Antanas Verikas</i>	
Active Learning Strategies: A Case Study for Detection of Emotions in Speech	228
<i>Alexis Bondu, Vincent Lemaire, and Barbara Poulain</i>	
Neural Business Control System	242
<i>Maria Lourdes Borrajo, Juan M. Corchado, Emilio S. Corchado, and M. Asumptio Pellicer</i>	
A Framework for Discovering and Analyzing Changing Customer Segments	255
<i>Mirko Böttcher, Martin Spott, and Detlef Nauck</i>	
Collaborative Filtering Using Electrical Resistance Network Models	269
<i>Jérôme Kunegis and Stephan Schmidt</i>	
Visual Query and Exploration System for Temporal Relational Database	283
<i>Shaul Ben Michael and Ronen Feldman</i>	

Towards an Online Image-Based Tree Taxonomy	296
<i>Paul M. de Zeeuw, Elena Rangelova, and Eric J. Pauwels</i>	
Distributed Generative Data Mining	307
<i>Ruy Ramos and Rui Camacho</i>	

Time Series and Frequent Pattern Mining

Privacy-Preserving Discovery of Frequent Patterns in Time Series	318
<i>Josenildo Costa da Silva and Matthias Klusch</i>	
Efficient Non Linear Time Series Prediction Using Non Linear Signal Analysis and Neural Networks in Chaotic Diode Resonator Circuits	329
<i>Michalis P. Haniias and Dimitrios A. Karras</i>	

Association Mining

Using Disjunctions in Association Mining	339
<i>Martin Ralbovský and Tomáš Kuchař</i>	

Author Index	353
------------------------	-----