Lecture Notes in Artificial Intelligence 4265

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Ljupčo Todorovski Nada Lavrač
Klaus P. Jantke (Eds.)

Discovery Science

9th International Conference, DS 2006
Barcelona, Spain, October 7-10, 2006
Proceedings

Springer

Series Editors

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

Volume Editors

Ljupčo Todorovski
Nada Lavrač
Jožef Stefan Institute
Department of Knowledge Technologies
Jamova 39, 1000 Ljubljana, Slovenia
E-mail:{ljupco.todorovski, nada.lavrac}@ijs.si

Klaus P. Jantke
Technical University Ilmenau
Institut für Medien- und Kommunikationswissenschaft
PF 10 05 65, 98684 Ilmenau, Germany
E-mail: klaus-peter.jantke@tu-ilmenau.de

Library of Congress Control Number: 2006933944

CR Subject Classification (1998): I.2, H.2.8, H.3, J.1, J.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-46491-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-46491-4 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: 11893318 06/3142 5 4 3 2 1 0

Preface

The 9th International Conference on Discovery Science (DS 2006) was held in Barcelona, Spain, on 7–10 October 2006. The conference was collocated with the 17th International Conference on Algorithmic Learning Theory (ALT 2006). The two conferences shared the invited talks.

This LNAI volume, containing the proceedings of the 9th International Conference on Discovery Science, is structured in three parts. The first part contains the papers/abstracts of the invited talks, the second part contains the accepted long papers, and the third part the accepted regular (short) papers. Out of 87 submitted papers, 23 were accepted for publication as long papers, and 18 as regular papers. All the submitted papers were reviewed by two or three referees. In addition to the presentations of accepted papers, the DS 2006 conference program consisted of three invited talks, two tutorials, the collocated ALT 2006 conference and the Pascal Dialogues workshop.

We wish to express our gratitude to

- the authors of submitted papers,
- the program committee and other referees for their thorough and timely paper evaluation,
- DS 2006 invited speakers Carole Goble and Padhraic Smyth, as well as Andrew Ng as joint DS 2006 and ALT 2006 invited speaker,
- invited tutorial speakers Luis Torgo and Michael May,
- the local organization committee chaired by Ricard Gavaldà,
- DS 2006 conference chair Klaus P. Jantke,
- the DS steering committee, chaired by Hiroshi Motoda,
- Andrei Voronkov for the development of EasyChair which provided excellent support in the paper submission, evaluation and proceedings production process,
- Alfred Hofmann of Springer for the co-operation in publishing the proceedings,
- the ALT 2006 PC chairs Phil Long and Frank Stephan, as well as Thomas Zeugman and José L. Balcázar, for the cooperation and coordination of the two conferences, and finally
- we gratefully acknowledge the financial support of the Universitat Politècnica de Catalunya, Idescat — the Statistical Institute of Catalonia (for providing support to tutorial speakers), the PASCAL Network of Excellence (for supporting the Pascal Dialogues), the Spanish Ministry of Science and Education, the Slovenian Ministry of Higher Education, Science, and Technology, and Yahoo! Research for sponsoring the Carl Smith Student Award.

We hope that the week in Barcelona in early October 2006 was a fruitful, challenging and enjoyable scientific and social event.

August 2006 Nada Lavrač and Ljupčo Todorovski

Conference Organization

Conference Chair

Klaus Jantke	Technical University of Ilmenau, Germany

Steering Committee Chair

Hiroshi Motoda	AFOSR/AOARD & Osaka University, Japan

Program Committee Chairs

Nada Lavrač	Jožef Stefan Institute, Slovenia
Ljupčo Todorovski	Jožef Stefan Institute, Slovenia

Program Committee

José Luis Balcázar	Universitat Politècnica de Catalunya, Spain
Michael Berthold	University of Konstanz, Germany
Elisa Bertino	Purdue University, USA
Vincent Corruble	Pierre and Marie Curie University, France
Andreas Dress	Shanghai Institutes for Biological Sciences, China
Sašo Džeroski	Jožef Stefan Institute, Slovenia
Tapio Elomaa	Tampere University of Technology, Finland
João Gama	University of Porto, Portugal
Dragan Gamberger	Rudjer Bošković Institute, Croatia
Gunter Grieser	Technical University of Darmstadt, Germany
Fabrice Guillet	University of Nantes, France
Mohand-Saïd Hacid	Claude Bernard University Lyon 1, France
Udo Hahn	Jena University, Germany
Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Achim Hoffmann	University of New South Wales, Australia
Szymon Jaroszewicz	Szczecin University of Technology, Poland
Kristian Kersting	University of Freiburg, Germany
Ross King	University of Wales, UK
Kevin Korb	Monash University, Australia
Ramamohanarao Kotagiri	University of Melbourne, Australia
Stefan Kramer	Technical University of Munich, Germany
Nicolas Lachiche	Louis Pasteur University, France
Aleksandar Lazarević	University of Minnesota, USA
Jinyan Li	Institute for Infocomm Research, Singapore

Ashesh Mahidadia	University of New South Wales, Australia
Michael May	Fraunhofer Institute for Autonomous Intelligent Systems, Germany
Dunja Mladenić	Jožef Stefan Institute, Slovenia
Igor Mozetič	Jožef Stefan Institute, Slovenia
Ion Muslea	Language Weaver, USA
Lourdes Peña Castillo	University of Toronto, Canada
Bernhard Pfahringer	University of Waikato, New Zealand
Jan Rauch	University of Economics, Prague, Czech Republic
Domenico Saccà	University of Calabria, Italy
Rudy Setiono	National University of Singapore, Singapore
Einoshin Suzuki	Kyushu University, Japan
Masayuki Takeda	Kyushu Univeristy, Japan
Kai Ming Ting	Monash University, Australia
Alfonso Valencia	National Centre for Biotechnology, Spain
Takashi Washio	Osaka University, Japan
Gerhard Widmer	Johannes Kepler University, Austria
Akihiro Yamamoto	Kyoto University, Japan
Mohammed Zaki	Rensselaer Polytechnic Institute, USA
Filip Železný	Czech Technical University, Czech Republic
Djamel A. Zighed	Lumière University Lyon 2, France

Local Organization

Ricard Gavaldà (Chair)	Universitat Politècnica de Catalunya, Spain
José Luis Balcázar	Universitat Politècnica de Catalunya, Spain
Albert Bifet	Universitat Politècnica de Catalunya, Spain
Gemma Casas	Universitat Politècnica de Catalunya, Spain
Jorge Castro	Universitat Politècnica de Catalunya, Spain
Jesús Cerquides	Universitat de Barcelona, Spain
Pedro Delicado	Universitat Politècnica de Catalunya, Spain
Gábor Lugosi	Universitat Pompeu Fabra, Spain
Victor Dalmau	Universitat Pompeu Fabra, Spain

External Reviewers

Zeyar Aung
Jose Campos Avila
Hideo Bannai
Maurice Bernadet
Julien Blanchard
Bruno Bouzy
Janez Brank
Agnès Braud
Jérôme David
Stéphane Daviet
Fabian Dill
Tomaž Erjavec
Jure Ferlež
Blaž Fortuna
Gemma C. Garriga
Gianluigi Greco
El Ghazel Haytham
Hakim Hacid
Kohei Hatano
Le Minh Hoang
Zujun Hou
Qingmao Hu
Salvatore Iritano
Daisuke Ikeda
Ignazio Infantino
Aneta Ivanovska
Jussi Kujala
Ha Thanh Le
Aydano Machado
Giuseppe Manco
Blaž Novak
Riccardo Ortale
François Pachet
Kwok Pan Pang
Panče Panov
Nguyen Thanh Phuong
Gerard Ramstein
Lothar Richter
Pedro Rodrigues
Massimo Ruffolo
Ulrich Rückert
Ivica Slavkov
James Tan Swee-Chuan
Miha Vuk
Bernd Wiswedel
Bernard Ženko

Table of Contents

I Invited Papers

e-Science and the Semantic Web: A Symbiotic Relationship 1
Carole Goble, Oscar Corcho, Pinar Alper, David De Roure

Data-Driven Discovery Using Probabilistic Hidden Variable Models 13
Padhraic Smyth

Reinforcement Learning and Apprenticeship Learning for Robotic Control 14
Andrew Ng

The Solution of Semi-Infinite Linear Programs Using Boosting-Like Methods 15
Gunnar Rätsch

Spectral Norm in Learning Theory: Some Selected Topics 16
Hans Ulrich Simon

II Long Papers

Classification of Changing Regions Based on Temporal Context in Local Spatial Association 17
Jae-Seong Ahn, Yang-Won Lee, Key-Ho Park

Kalman Filters and Adaptive Windows for Learning in Data Streams 29
Albert Bifet, Ricard Gavaldà

Scientific Discovery: A View from the Trenches 41
Catherine Blake, Meredith Rendall

Optimal Bayesian 2D-Discretization for Variable Ranking in Regression 53
Marc Boullé, Carine Hue

Text Data Clustering by Contextual Graphs 65
Krzysztof Ciesielski, Mieczysław A. Kłopotek

Automatic Water Eddy Detection in SST Maps Using Random Ellipse Fitting and Vectorial Fields for Image Segmentation 77
Armando Fernandes, Susana Nascimento

Mining Approximate Motifs in Time Series 89
Pedro G. Ferreira, Paulo J. Azevedo, Cândida G. Silva, Rui M.M. Brito

Identifying Historical Period and Ethnic Origin of Documents Using Stylistic Feature Sets 102
Yaakov HaCohen-Kerner, Hananya Beck, Elchai Yehudai, Dror Mughaz

A New Family of String Classifiers Based on Local Relatedness 114
Yasuto Higa, Shunsuke Inenaga, Hideo Bannai, Masayuki Takeda

On Class Visualisation for High Dimensional Data: Exploring Scientific Data Sets 125
Ata Kabán, Jianyong Sun, Somak Raychaudhury, Louisa Nolan

Mining Sectorial Episodes from Event Sequences 137
Takashi Katoh, Kouichi Hirata, Masateru Harao

A Voronoi Diagram Approach to Autonomous Clustering 149
Heidi Koivistoinen, Minna Ruuska, Tapio Elomaa

Itemset Support Queries Using Frequent Itemsets and Their Condensed Representations 161
Taneli Mielikäinen, Panče Panov, Sašo Džeroski

Strategy Diagram for Identifying Play Strategies in Multi-view Soccer Video Data 173
Yukihiro Nakamura, Shin Ando, Kenji Aoki, Hiroyuki Mano, Einoshin Suzuki

Prediction of Domain-Domain Interactions Using Inductive Logic Programming from Multiple Genome Databases 185
Thanh Phuong Nguyen, Tu Bao Ho

Clustering Pairwise Distances with Missing Data: Maximum Cuts Versus Normalized Cuts 197
Jan Poland, Thomas Zeugmann

Analysis of Linux Evolution Using Aligned Source Code Segments 209
Antti Rasinen, Jaakko Hollmén, Heikki Mannila

Rule-Based Prediction of Rare Extreme Values 219
Rita Ribeiro, Luís Torgo

A Pragmatic Logic of Scientific Discovery 231
Jean Sallantin, Christopher Dartnell, Mohammad Afshar

Change Detection with Kalman Filter and CUSUM 243
Milton Severo, João Gama

Automatic Recognition of Landforms on Mars Using Terrain Segmentation and Classification 255
Tomasz F. Stepinski, Soumya Ghosh, Ricardo Vilalta

A Multilingual Named Entity Recognition System Using Boosting and C4.5 Decision Tree Learning Algorithms 267
György Szarvas, Richárd Farkas, András Kocsor

Model-Based Estimation of Word Saliency in Text 279
Xin Wang, Ata Kabán

III Regular Papers

Learning Bayesian Network Equivalence Classes from Incomplete Data 291
Hanen Borchani, Nahla Ben Amor, Khaled Mellouli

Interesting Patterns Extraction Using Prior Knowledge 296
Laurent Brisson

Visual Interactive Subgroup Discovery with Numerical Properties of Interest 301
Alípio M. Jorge, Fernando Pereira, Paulo J. Azevedo

Contextual Ontological Concepts Extraction 306
Lobna Karoui, Nacéra Bennacer, Marie-Aude Aufaure

Experiences from a Socio-economic Application of Induction Trees 311
Fabio B. Losa, Pau Origoni, Gilbert Ritschard

Interpreting Microarray Experiments Via Co-expressed Gene Groups Analysis (CGGA) 316
Ricardo Martinez, Nicolas Pasquier, Claude Pasquier, Lucero Lopez-Perez

Symmetric Item Set Mining Based on Zero-Suppressed BDDs 321
Shin-ichi Minato

Mathematical Models of Category-Based Induction 327
Mizuho Mishima, Makoto Kikuchi

Automatic Construction of Static Evaluation Functions for Computer Game Players 332
Makoto Miwa, Daisaku Yokoyama, Takashi Chikayama

Databases Reduction Simultaneously by Ordered Projection 337
Isabel Nepomuceno, Juan A. Nepomuceno, Roberto Ruiz, Jesús S. Aguilar–Ruiz

Mapping Ontologies in an Air Pollution Monitoring and Control Agent-Based System 342
Mihaela Oprea

Information Theory and Classification Error in Probabilistic Classifiers 347
Aritz Pérez, Pedro Larrañaga, Iñaki Inza

Checking Scientific Assumptions by Modeling 352
Joseph Phillips, Ronald Edwards, Raghuveer Kumarakrishnan

Incremental Algorithm Driven by Error Margins 358
Gonzalo Ramos-Jiménez, José del Campo-Ávila, Rafael Morales-Bueno

Feature Construction and δ-Free Sets in 0/1 Samples 363
Nazha Selmaoui, Claire Leschi, Dominique Gay, Jean-François Boulicaut

Visual Knowledge Discovery in Paleoclimatology with Parallel Coordinates 368
Roberto Therón

A Novel Framework for Discovering Robust Cluster Results 373
Hye-Sung Yoon, Sang-Ho Lee, Sung-Bum Cho, Ju Han Kim

Gene Selection for Classifying Microarray Data Using Grey Relation Analysis 378
Li-Juan Zhang, Zhou-Jun Li

Author Index 383