

Lecture Notes in Artificial Intelligence 6332

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

Subseries of Lecture Notes in Computer Science

Bernhard Pfahringer Geoff Holmes
Achim Hoffmann (Eds.)

Discovery Science

13th International Conference, DS 2010
Canberra, Australia, October 6-8, 2010
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 Editors

Bernhard Pfahringer

Geoff Holmes

The University of Waikato, Department of Computer Science

Private Bag 3105, Hamilton 3240, New Zealand

E-mail: {bernhard, geoff}@cs.waikato.ac.nz

Achim Hoffmann

The University of New South Wales, School of Computer Science and Engineering

Sydney 2052, Australia

E-mail: achim@cse.unsw.edu.au

Library of Congress Control Number: 2010935451

CR Subject Classification (1998): I.2, H.3, H.4, H.2.8, J.1, F.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-16183-9 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-16183-4 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.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

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

Printed on acid-free paper 06/3180

Preface

This volume contains the papers presented at the 13th International Conference on Discovery Science (DS 2010) held in Canberra, Australia, October 6–8, 2010.

The main objective of the Discovery Science (DS) conference series is to provide an open forum for intensive discussions and the exchange of new ideas and information among researchers working in the area of automating scientific discovery or working on tools for supporting the human process of discovery in science. It has been a successful arrangement in the past to co-locate the DS conference with the International Conference on Algorithmic Learning Theory (ALT). This combination of ALT and DS allows for a comprehensive treatment of the whole range, from theoretical investigations to practical applications. Continuing in this tradition, DS 2010 was co-located with the 21st ALT conference (ALT 2010). The proceedings of ALT 2010 were published as a twin volume (6331) of the LNCS series.

The international steering committee of the Discovery Science conference series provided important advice on a number of issues during the planning of Discovery Science 2010. The members of the steering committee were Alberto Apostolico, Setsuo Arikawa, Hiroki Arimura, Jean-Francois Boulicaut, Vitor Santos Costa, Vincent Corruble, Joao Gama, Achim Hoffmann, Tamas Horvath, Alipio Jorge, Hiroshi Motoda, Ayumi Shinohara, Einoshin Suzuki (Chair), Masayuki Takeda, Akihiro Yamamoto, and Thomas Zeugmann.

We received 43 full-paper submissions out of which 25 long papers were accepted for presentation and are published in this volume. Each submission was allocated three reviewers from the program committee of international experts in the field. In total 125 reviews were written discussing in detail the merits of each submission. The selection of papers was made after careful evaluation of each paper based on originality, technical quality, relevance to the field of Discovery Science, and clarity.

This volume consists of two parts. The first part contains the papers accepted for presentation at the conference.

The second part contains the invited talks of ALT 2010 and DS 2010. Since the talks were shared between the two conferences, for the speakers invited specifically for ALT 2010, only abstracts are contained in this volume, while the full papers are found in the twin volume, LNCS 6331 (the proceedings of ALT 2010). The following invited speakers presented their work: Peter Bartlett *Optimal Online Prediction in Adversarial Environments*, Ivan Bratko *Discovery of Abstract Concepts by a Robot*, Alexander Clark, *Towards General Algorithms for Grammatical Inference*, Rao Kotagiri *Contrast Pattern Mining and Its Application for Building Robust Classifiers*, and Manfred Warmuth *The Blessing and the Curse of the Multiplicative Updates*.

We are deeply indebted to the program committee members as well as their subreferees who had the critically important role of reviewing the submitted papers and contributing to the intense discussions which resulted in the selection of the papers published in this volume. Without this enormous effort, ensuring the high quality of the work presented at Discovery Science 2010 would not have been possible.

We also thank all the authors who submitted their work to Discovery Science 2010 for their efforts. We wish to express our gratitude to the invited speakers for their acceptance of the invitation and their stimulating contributions to the conference. Furthermore, we wish to thank the *Air Force Office of Scientific Research*, *Asian Office of Aerospace Research and Development* as well as the *Artificial Intelligence Journal* for their financial support contributing to the success of this conference.

Finally, we wish to thank everyone who helped to make Discovery Science 2010 a success: the DS steering committee, the ALT conference chairs, invited speakers, the Publicity Chair for Discovery Science 2010, Albert Bifet, for the well-designed web presence, and last but not least the Local Arrangements Chair for Discovery Science 2010, Eric McCreath, and the Local Arrangements Chair for Algorithmic Learning Theory, Mark Reid, and their team of supporters who worked very hard to make both conferences a success.

July 2010

Bernhard Pfahringer
Geoff Holmes
Achim Hoffmann

Organization

Organization Committee

Steering Committee Chair	Einoshin Suzuki
Conference Chair	Achim Hoffmann
Program Chairs	Bernhard Pfahringer Geoffrey Holmes
Publicity Chair	Albert Bifet
Local Arrangements Chair	Eric McCreath

Program Committee

Akihiro Yamamoto	Kyoto University, Japan
Albert Bifet	University of Waikato, New Zealand
Albert Yeap	AUT, New Zealand
Alipio Jorge	University of Porto, Portugal
Alneu de Andrade Lopes	University of Sao Paulo, Brazil
Andre Carvalho	University of Sao Paulo, Brazil
Antoine Cornuejols	AgroParisTech, France
Antonio Bahamonde	University of Oviedo, Spain
Bettina Berendt	Katholieke Universiteit Leuven, Belgium
Carlos Soares	University of Porto, Portugal
Colin de la Higuera	University of Nantes, France
Concha Bielza	Technical University of Madrid, Spain
Daisuke Ikeda	Kyushu University, Japan
Daniel Berrar	University of Ulster, Ireland
David Dowe	Monash University, Melbourne, Australia
Dino Pedreschi	Pisa University, Italy
Donato Malerba	University of Bari, Italy
Einoshin Suzuki	Kyushu University, Japan
Filip Zelezny	Technical University Prague, Czech Republic
Gerhard Widmer	Johannes Kepler University, Austria
Gerson Zaverucha	Universidade Federal do Rio de Janeiro, Brazil
Guillaume Beslon	INSA Lyon, France
Hendrik Blockeel	Katholieke Universiteit Leuven, Belgium
Hideo Bannai	Kyushu University, Fukuoka, Japan
Hiroshi Motoda	AOARD and Osaka University, Japan
Hisashi Kashima	University of Tokyo, Japan
Ian Watson	University of Auckland, New Zealand
Inaki Inza	University of Basque Country, Spain
Ingrid Fischer	University of Konstanz, Germany
Irene Ong	UW-Madison, USA

VIII Organization

Jaakko Hollmen	Helsinki University of Technology, Finland
Janos Csirik	University of Szeged and RGAI Szeged, Hungary
Jean-Francois Boulicaut	INSA Lyon, France
Jesse Davis	University of Washington, USA
Joao Gama	University of Porto, Portugal
Jose Luis Borges	University of Porto, Portugal
Kevin Korb	Monash University, Australia
Kouichi Hirata	Kyushu Institute of Technology, Japan
Kristian Kersting	Fraunhofer IAIS and University of Bonn, Germany
Kuniaki Uehara	Kobe University, Japan
Ljupco Todorovski	Josef Stefan Institute, Slovenia
Luis Torgo	University of Porto, Portugal
Maarten van Someren	University of Amsterdam, Netherlands
Maguelonne Teisseire	Cemagraf Montpellier, France
Makoto Haraguchi	Hokkaido University, Japan
Marta Arias	Universitat Politècnica de Catalunya, Spain
Mario J. Silva	Universidade de Lisboa, Portugal
Masayuki Numao	University of Osaka, Japan
Michael Berthold	University of Konstanz, Germany
Michael May	Fraunhofer IAIS Bonn, Germany
Mohand-Said Hacid	University Claude Bernard Lyon 1, France
Nada Lavrač	Jožef Stefan Institute, Slovenia
Nuno Fonseca	University of Porto, Portugal
Patricia Riddle	University of Auckland, New Zealand
Paulo Azevedo	Universidade do Minho, Portugal
Pedro Larranaga	Polytechnic University of Madrid, Spain
Peter Andreae	Victoria University of Wellington, New Zealand
Peter Christen	Australian National University, Australia
Ross King	University of Wales, UK
Simon Colton	Imperial College London, UK
Sriraam Natarajan	UW-Madison, USA
Stan Matwin	University Ottawa, Canada
Stefan Kramer	Technische Universität München, Germany
Stephen Marsland	Massey University, New Zealand
Szymon Jaroszewicz	National Institute of Telecommunications, Poland
Takashi Washio	Osaka University, Japan
Tamas Horvath	University of Bonn and Fraunhofer IAIS, Germany
Tapio Elomaa	Tampere University of Technology, Finland
Vincent Corruble	Université Pierre et Marie Curie, France
Vitor Santos Costa	University of Porto, Portugal
Will Bridewell	CSLI Stanford, USA

Additional Reviewers

Carlos Ferreira
Cristiano Pitangui

Elena Ikonomovska
Ken-ichi Fukui

Table of Contents

Sentiment Knowledge Discovery in Twitter Streaming Data	1
<i>Albert Bifet and Eibe Frank</i>	
A Similarity-Based Adaptation of Naive Bayes for Label Ranking: Application to the Metalearning Problem of Algorithm Recommendation	16
<i>Artur Aiguzhinov, Carlos Soares, and Ana Paula Serra</i>	
Topology Preserving SOM with Transductive Confidence Machine	27
<i>Bin Tong, ZhiGuang Qin, and Eimoshin Suzuki</i>	
An Artificial Experimenter for Enzymatic Response Characterisation . . .	42
<i>Chris Lovell, Gareth Jones, Steve R. Gunn, and Klaus-Peter Zauner</i>	
Subgroup Discovery for Election Analysis: A Case Study in Descriptive Data Mining	57
<i>Henrik Grosskreutz, Mario Boley, and Maike Krause-Traudes</i>	
On Enumerating Frequent Closed Patterns with Key in Multi-relational Data	72
<i>Hirohisa Seki, Yuya Honda, and Shinya Nagano</i>	
Why Text Segment Classification Based on Part of Speech Feature Selection	87
<i>Iulia Nagy, Katsuyuki Tanaka, and Yasuo Ariki</i>	
Speeding Up and Boosting Diverse Density Learning	102
<i>James R. Foulds and Eibe Frank</i>	
Incremental Learning of Cellular Automata for Parallel Recognition of Formal Languages	117
<i>Katsuhiko Nakamura and Keita Imada</i>	
Sparse Substring Pattern Set Discovery Using Linear Programming Boosting	132
<i>Kazuaki Kashiwara, Kohei Hatano, Hideo Bannai, and Masayuki Takeda</i>	
Discovery of Super-Mediators of Information Diffusion in Social Networks	144
<i>Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda</i>	
Integer Linear Programming Models for Constrained Clustering	159
<i>Marianne Mueller and Stefan Kramer</i>	

Efficient Visualization of Document Streams	174
<i>Miha Grčar, Vid Podpečan, Matjaž Jursič, and Nada Lavrač</i>	
Bridging Conjunctive and Disjunctive Search Spaces for Mining a New Concise and Exact Representation of Correlated Patterns	189
<i>Nassima Ben Younes, Tarek Hamrouni, and Sadok Ben Yahia</i>	
Graph Classification Based on Optimizing Graph Spectra	205
<i>Nguyen Duy Vinh, Akihiro Inokuchi, and Takashi Washio</i>	
Algorithm for Detecting Significant Locations from Raw GPS Data	221
<i>Nobuharu Kami, Nobuyuki Enomoto, Teruyuki Baba, and Takashi Yoshikawa</i>	
Discovery of Conservation Laws via Matrix Search	236
<i>Oliver Schulte and Mark S. Drew</i>	
Gaussian Clusters and Noise: An Approach Based on the Minimum Description Length Principle	251
<i>Panu Luosto, Jyrki Kivinen, and Heikki Mannila</i>	
Exploiting Code Redundancies in ECOC	266
<i>Sang-Hyeun Park, Lorenz Weizsäcker, and Johannes Fürnkranz</i>	
Concept Convergence in Empirical Domains	281
<i>Santiago Ontañón and Enric Plaza</i>	
Equation Discovery for Model Identification in Respiratory Mechanics of the Mechanically Ventilated Human Lung	296
<i>Steven Ganzert, Josef Guttman, Daniel Steinmann, and Stefan Kramer</i>	
Mining Class-Correlated Patterns for Sequence Labeling	311
<i>Thomas Hopf and Stefan Kramer</i>	
ESTATE: Strategy for Exploring Labeled Spatial Datasets Using Association Analysis	326
<i>Tomasz F. Stepinski, Josue Salazar, Wei Ding, and Denis White</i>	
Adapted Transfer of Distance Measures for Quantitative Structure-Activity Relationships	341
<i>Ulrich Rückert, Tobias Girschick, Fabian Buchwald, and Stefan Kramer</i>	
Incremental Mining of Closed Frequent Subtrees	356
<i>Viet Anh Nguyen and Akihiro Yamamoto</i>	
Optimal Online Prediction in Adversarial Environments	371
<i>Peter L. Bartlett</i>	

Discovery of Abstract Concepts by a Robot	372
<i>Ivan Bratko</i>	
Contrast Pattern Mining and Its Application for Building Robust Classifiers	380
<i>Kotagiri Ramamohanarao</i>	
Towards General Algorithms for Grammatical Inference.....	381
<i>Alexander Clark</i>	
The Blessing and the Curse of the Multiplicative Updates	382
<i>Manfred K. Warmuth</i>	
Author Index	383