

Lecture Notes in Artificial Intelligence 6926

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

Tapio Elomaa Jaakko Hollmén
Heikki Mannila (Eds.)

Discovery Science

14th International Conference, DS 2011
Espoo, Finland, October 5-7, 2011
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

Tapio Elomaa
Tampere University of Technology, Department of Software Systems
P.O. Box 553, 33101 Tampere, Finland
E-mail: elomaa@cs.tut.fi

Jaakko Hollmén
Heikki Mannila
Aalto University School of Science
Department of Information and Computer Science
P.O. Box 15400, 00076 Aalto, Finland
E-mail: {jaakko.hollmen, heikki.mannila}@aalto.fi

ISSN 0302-9743
ISBN 978-3-642-24476-6
DOI 10.1007/978-3-642-24477-3
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-24477-3

Library of Congress Control Number: 2011937030

CR Subject Classification (1998): I.2, H.3, H.4, H.2.8, J.1, H.5, C.2

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

We are pleased to present the proceedings of the 14th International Conference on Discovery Science (DS 2011), held in Espoo, Finland, during October 5-7, 2011. DS 2011 provided an open forum for intensive discussions and exchange of new ideas among researchers working in the area of discovery science. The scope of the conference includes the development and analysis of methods for automatic scientific knowledge discovery, machine learning, intelligent data analysis, theory of learning, as well as their application to knowledge discovery. DS 2011 was co-located with the 22nd International Conference on Algorithmic Learning Theory (ALT 2011). The proceedings of ALT 2011 appeared as companion volume number 6925 in the Springer LNCS series. This year we had a decennial celebration of the fruitful co-location of the ALT and DS conference series, which allowed us to cover the whole range from theoretical explorations to practical applications.

We were honored to have a strong program. We received 56 submissions from authors coming from 22 countries. The submissions were reviewed by three members of the Program Committee; the Program Chairs evaluated the review reports and decided to accept 24 papers based on the merits of originality, technical quality, and relevance to the field of discovery science. The overall acceptance rate was 43%. We were delighted to have five prestigious invited talks by six speakers. Ming Li from the University of Waterloo in Canada presented “Information Distance and Its Extensions”. Eyke Hüllermeier from Philipps-Universität Marburg in Germany together with Johannes Fürnkranz from Technische Universität Darmstadt in Germany presented a joint invited talk on the latest research concerning “Learning from Label Preferences”. Yoshua Bengio from the University of Montreal in Canada was the joint invited speaker of ALT 2011 and DS 2011, and he presented “On the Expressive Power of Deep Architectures”. Jorma Rissanen from Helsinki Institute for Information Technology in Finland presented an invited talk on “Optimal Estimation”. Peter Auer from the University of Leoben in Austria presented “Models for Autonomously Motivated Exploration in Reinforcement Learning”. Tutorials were given on the topic of preference learning by Eyke Hüllermeier and Johannes Fürnkranz jointly and on the topic of bandits by Sébastien Bubeck from Princeton University in USA.

We wish to express our gratitude to all authors of submitted papers, the Program Committee members and the additional reviewers for their efforts in evaluating the submitted papers, as well as the invited speakers and tutorial presenters. We thank the members of the DS Steering Committee, especially its Chair, Einoshin Suzuki, who were very helpful in providing continual support necessary for the continuity of the DS series. We thank the ALT editors Jyrki Kivinen, Csaba Szepesvári, Esko Ukkonen, and Thomas Zeugmann for the fruitful coordination of the co-located ALT and DS conferences. We also thank

the Local Organizing Committee, chaired by Olli Simula, for making the practical arrangements concerning the ALT and DS conferences. We acknowledge the support from our hosting organizations Aalto University, University of Helsinki, Algodan - Centre of Excellence for Algorithmic Data Analysis Research, and Helsinki Institute for Information Technology in Finland. Finally, we thank the *Artificial Intelligence* journal for the financial support.

August 2011

Tapio Elomaa
Jaakko Hollmén
Heikki Mannila

Organization

Organizing Committee

Steering Committee Chair

Einoshin Suzuki

Conference Chair

Heikki Mannila

Program Chairs

Tapio Elomaa

Jaakko Hollmén

Local Organization Chair

Olli Simula

Local Organizing Committee

Ella Bingham

Tapio Elomaa

Jaakko Hollmén

Jyrki Kivinen

Mikko Korpela

Kai Puolamäki

Mika Sulkava

Janne Toivola

Program Committee

Hiroki Arimura

Hideo Bannai

Michael R. Berthold

Hendrik Blockeel

Henrik Bostrom

Jean-Francois Boulicaut

Bruno Crémilleux

Sašo Džeroski

Tapio Elomaa

Peter Flach

Johannes Fürnkranz

Hokkaido University, Japan

Kyushu University, Japan

University of Konstanz, Germany

K.U. Leuven, Belgium

Stockholm University, Sweden

INSA Lyon, France

University of Caen, France

Jožef Stefan Institute, Slovenia

Tampere University of Technology, Finland

University of Bristol, UK

TU Darmstadt, Germany

João Gama	Universidade do Porto, Portugal
Mohand-Said Hacid	Université Claude Bernard Lyon 1, France
Howard J. Hamilton	University of Regina, Canada
Makoto Haraguchi	Hokkaido University, Japan
Kouichi Hirata	Kyushu Institute of Technology, Japan
Jaakko Hollmén	Aalto University, Finland
Geoff Holmes	University of Waikato, New Zealand
Vasant Honavar	Iowa State University, USA
Tamás Horváth	University of Bonn, Fraunhofer IAIS, Germany
Eyke Hüllermeier	Philipps-Universität Marburg, Germany
Alípio M. Jorge	Universidade do Porto, Portugal
Hisashi Kashima	University of Tokyo, Japan
Ross King	University of Wales, Aberystwyth, UK
Joost N. Kok	LIACS, Leiden University, The Netherlands
Stefan Kramer	TU München, Germany
Nada Lavrač	Jožef Stefan Institute, Slovenia
Philippe Lenca	Télécom Bretagne, France
Donato Malerba	Università di Bari, Italy
Taneli Mielikäinen	Nokia Research Center, USA
Richard Nock	Université des Antilles et de la Guyane, France
Panagiotis Papapetrou	Aalto University, Finland
Mykola Pechenizkiy	Eindhoven University of Technology, The Netherlands
Jean-Marc Petit	University of Lyon, INSA Lyon, France
Bernhard Pfahringer	University of Waikato, New Zealand
Enric Plaza	IIIA-CSIC, Spain
Juho Rousu	University of Helsinki, Finland
Jean Sallantin	CNRS, France
Vítor Santos Costa	Universidade do Porto, Portugal
Mika Sulkava	Aalto University, Finland
Einoshin Suzuki	Kyushu University, Japan
Nikolaj Tatti	University of Antwerp, Belgium
Hannu Toivonen	University of Helsinki, Finland
Antti Ukkonen	Yahoo! Research, Spain
Akihiro Yamamoto	Kyoto University, Japan
Zhi-Hua Zhou	Nanjing University, China

Additional Reviewers

Ceci, Michelangelo	Koskela, Markus
De Knijf, Jeroen	Kuboyama, Tetsuji
Fonseca, Nuno	Li, Nan
Girschick, Tobias	Loglisci, Corrado
Heinimäki, Teemu	Oliveira, Marcia
Kosina, Petr	Ontañón, Santiago

Plantevit, Marc
Sakamoto, Hiroshi
Schietgat, Leander
Schmidt, Jana
Schmitt, Constanze
Seeland, Madeleine

Slavkov, Ivica
Sluban, Borut
Tashkova, Katerina
Vreeken, Jilles
Zenko, Bernard
Znidarsic, Martin

Table of Contents

On the Expressive Power of Deep Architectures	1
<i>Yoshua Bengio and Olivier Delalleau</i>	
Learning from Label Preferences	2
<i>Eyke Hüllermeier and Johannes Fürnkranz</i>	
Information Distance and Its Extensions	18
<i>Ming Li</i>	
Models for Autonomously Motivated Exploration in Reinforcement Learning	29
<i>Peter Auer, Shiao Hong Lim, and Chris Watkins</i>	
Optimal Estimation	30
<i>Jorma Rissanen</i>	
Monotone Instance Ranking with MIRA	31
<i>Nicola Barile and Ad Feelders</i>	
<i>MOA-TweetReader: Real-Time Analysis in Twitter Streaming Data</i>	46
<i>Albert Bifet, Geoffrey Holmes, and Bernhard Pfahringer</i>	
Application of Semantic Kernels to Literature-Based Gene Function Annotation	61
<i>Mathieu Blondel, Kazuhiro Seki, and Kuniaki Uehara</i>	
“Tell Me More”: Finding Related Items from User Provided Feedback. . .	76
<i>Jeroen De Knijff, Anthony Liekens, and Bart Goethals</i>	
MEI: Mutual Enhanced Infinite Generative Model for Simultaneous Community and Topic Detection	91
<i>Dongsheng Duan, Yuhua Li, Ruixuan Li, Zhengding Lu, and Aiming Wen</i>	
A Methodology for Mining Document-Enriched Heterogeneous Information Networks	107
<i>Miha Grčar and Nada Lavrač</i>	
Multiple Hypothesis Testing in Pattern Discovery	122
<i>Sami Hanhijärvi</i>	
A Parameter-Free Method for Discovering Generalized Clusters in a Network	135
<i>Hiroshi Hirai, Bin-Hui Chou, and Einoshin Suzuki</i>	

Detecting Anti-majority Opinionists Using Value-Weighted Mixture Voter Model	150
<i>Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda</i>	
Using Ontologies in Semantic Data Mining with SEGS and g-SEGS	165
<i>Nada Lavrač, Anže Vavpetič, Larisa Soldatova, Igor Trajkovski, and Petra Kralj Novak</i>	
Mining Classification Rules without Support: an Anti-monotone Property of Jaccard Measure	179
<i>Yannick Le Bras, Philippe Lenca, and Stéphane Lallich</i>	
Bootstrapping Parameter Estimation in Dynamic Systems	194
<i>Huma Lodhi and David Gilbert</i>	
Network Effects on Tweeting	209
<i>Jake T. Lussier and Nitesh V. Chawla</i>	
Context-Aware Personal Route Recognition	221
<i>Oleksiy Mazhelis, Indrė Žliobaitė, and Mykola Pechenzkiy</i>	
Scalable Detection of Frequent Substrings by Grammar-Based Compression	236
<i>Masaya Nakahara, Shirou Maruyama, Tetsuji Kuboyama, and Hiroshi Sakamoto</i>	
A Statistical Model for Topically Segmented Documents	247
<i>Giovanni Ponti, Andrea Tagarelli, and George Karypis</i>	
Predicting Structured Outputs k -Nearest Neighbours Method	262
<i>Mitja Pugelj and Sašo Džeroski</i>	
The Augmented Itemset Tree: A Data Structure for Online Maximum Frequent Pattern Mining	277
<i>Jana Schmidt and Stefan Kramer</i>	
Word Clouds for Efficient Document Labeling	292
<i>Christin Seifert, Eva Ulbrich, and Michael Granitzer</i>	
Global and Local Spatial Autocorrelation in Predictive Clustering Trees	307
<i>Daniela Stojanova, Michelangelo Ceci, Annalisa Appice, Donato Malerba, and Sašo Džeroski</i>	
Rule Stacking: An Approach for Compressing an Ensemble of Rule Sets into a Single Classifier	323
<i>Jan-Nikolas Sulzmann and Johannes Fürnkranz</i>	
Graph Clustering Based on Optimization of a Macroscopic Structure of Clusters	335
<i>Yuta Taniguchi and Daisuke Ikeda</i>	

Modeling the Temperature of Hot Rolled Steel Plate with Semi-supervised Learning Methods	351
<i>Henna Tiensuu, Ilmari Juutilainen, and Juha Rönning</i>	
Controlled Permutations for Testing Adaptive Classifiers	365
<i>Indrė Žliobaitė</i>	
Author Index	381