

Lecture Notes in Artificial Intelligence 4755

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Vincent Corruble Masayuki Takeda  
Einoshin Suzuki (Eds.)

# Discovery Science

10th International Conference, DS 2007  
Sendai, Japan, October 1-4, 2007  
Proceedings



Springer

## Series Editors

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

## Volume Editors

Vincent Corruble  
Université Pierre et Marie Curie (Paris 6)  
Laboratoire d'Informatique de Paris 6  
104 avenue du Président Kennedy, 75016 Paris, France  
E-mail: Vincent.Corruble@lip6.fr

Masayuki Takeda  
Einoshin Suzuki  
Kyushu University  
Department of Informatics  
744 Motooka, Nishi, Fukuoka 819-0395, Japan  
E-mail: {takeda, suzuki}@i.kyushu-u.ac.jp

Library of Congress Control Number: 2007935939

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 3-540-75487-3 Springer Berlin Heidelberg New York  
ISBN-13 978-3-540-75487-9 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: 12170210 06/3180 5 4 3 2 1 0

# Preface

This volume contains the papers presented at DS-2007: The Tenth International Conference on Discovery Science held in Sendai, Japan, October 1–4, 2007.

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 this tradition, DS 2007 was co-located with the 18th ALT conference (ALT 2007). The proceedings of ALT 2007 were published as a twin volume 4754 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 2007. The members of the Steering Committee are Einoshin Suzuki (Kyushu University, Chair), Achim G. Hoffmann (University of New South Wales, Vice Chair), Setsuo Arikawa (Kyushu University), Hiroshi Motoda (Osaka University), Masahiko Sato (Kyoto University), Satoru Miyano (University of Tokyo), Thomas Zeugmann (Hokkaido University), Ayumi Shinohara (Tohoku University), Alberto Apostolico (Georgia Institute of Technology and University of Padova), Massimo Melucci (University of Padova), Tobias Scheffer (Max Planck Institute for Computer Science), Ken Satoh (National Institute of Informatics), Nada Lavrac (Jozef Stefan Institute), Ljupco Todorovski (University of Ljubljana), and Hiroki Arimura (Hokkaido University).

In response to the call for papers 55 manuscripts were submitted. The Program Committee selected for publication 17 submissions as long papers and 10 submissions as regular papers. Each submission was reviewed by at least two members of the Program Committee, which consisted of international experts in the field. The selection was made after careful evaluation of each paper based on clarity, significance, technical quality, and originality, as well as relevance to the field of discovery science. This volume consists of 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 contains the accepted regular papers.

We are deeply indebted to the Program Committee members as well as their subreferees who played 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 their enormous effort, ensuring the high quality of the work presented at Discovery Science 2007 would not have been possible. Furthermore, we would like to thank all individuals and

institutions who contributed to the success of the conference: the authors for submitting papers, the invited speakers for their acceptance of the invitation and their stimulating contributions to the conference, the Steering Committee, and the sponsors for their support. In particular, we acknowledge the generous financial support from the Air Force Office of Scientific Research (AFOSR), Asian Office of Aerospace Research and Development (AOARD)<sup>1</sup>; the Graduate School of Information Sciences (GSIS), Tohoku University for providing secretarial assistance and equipment; the Research Institute of Electrical Communication (RIEC), Tohoku University; New Horizons in Computing, MEXT Grant-in-Aid for Scientific Research on Priority Areas; and the Semi-Structured Data Mining Project, MEXT Grant-in-Aid for Specially Promoted Research.

July 2007

Vincent Corruble  
Masayuki Takeda  
Einoshin Suzuki

---

<sup>1</sup> AFOSR/AOARD support is not intended to express or imply endorsement by the U. S. Federal Government.

# Conference Organization

## Conference Chair

Ayumi Shinohara                      Tohoku University, Japan

## Program Committee

Vincent Corruble (Co-chair)	Université Pierre et Marie Curie, Paris, France
Masayuki Takeda (Co-chair)	Kyushu University, Japan
Jean-Francois Boulicaut	INSA Lyon, France
Will Bridewell	CSLI, Stanford, USA
Simon Colton	Imperial College London, UK
Antoine Cornuejols	Université Paris-Sud, France
Andreas Dress	Shanghai Institutes for Biological Sciences, China
Saso Dzeroski	Jozef Stefan Institute, Slovenia
Tapio Elomaa	Tampere University of Technology, Finland
Johannes Fuernkranz	Technical University of Darmstadt, Germany
Dragan Gamberger	Rudjer Boskovic Institute, Hungary
Ricard Gavalda	Technical University of Catalonia, Spain
Gunter Grieser	Technical University of Darmstadt, Germany
Fabrice Guillet	Ecole Polytechnique of the University of Nantes, France
Mohand-Said Hacid	Université Lyon 1, France
Udo Hahn	Jena University, Germany
Makoto Haraguchi	Hokkaido University, Japan
Tomoyuki Higuchi	The Institute of Statistical Mathematics, Japan
Kouichi Hirata	Kyushu Institute of Technology, Japan
Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Achim Hoffmann	University of New South Wales, Australia
Tamás Horváth	Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany
Daisuke Ikeda	Kyushu University, Japan
Kentaro Inui	Nara Institute of Science and Technology, Japan
Szymon Jaroszewicz	National Institute of Telecommunications, Poland
Hisashi Kashima	IBM Research, Tokyo Research Laboratory, Japan
Kristian Kersting	Universitaet Freiburg, Germany
Ross King	University of Wales, UK
Andras Kocsor	University of Szeged, Hungary
Kevin Korb	Monash University, Australia
Stefan Kramer	Technical University of Munich, Germany
Nicolas Lachiche	Université de Strasbourg, France

## VIII Organization

Nada Lavrac  
Aleksandar Lazarevic  
Jinyan Li  
Yuji Matsumoto  
Michael May

Taneli Mielikäinen  
Tetsuhiro Miyahara  
Dunja Mladenic  
Yasuhiko Morimoto  
Tsuyoshi Murata  
Bernhard Pfahringer  
Jan Rauch  
Kazumi Saito  
Michèle Sébag  
Rudy Setiono  
Einoshin Suzuki  
Kai Ming Ting  
Ljupco Todorovski  
Luis Torgo  
Kuniaki Uehara  
Takashi Washio  
Gerhard Widmer  
Akihiro Yamamoto  
Mohammed Zaki  
Filip Zelezny  
  
Jean-Daniel Zucker

Jozef Stefan Institute, Slovenia  
United Technologies Research Center, USA  
Institute for Infocomm Research, Singapore  
Nara Institute of Science and Technology, Japan  
Fraunhofer Institute for Autonomous Intelligent  
Systems, Germany  
Nokia Research Center, Palo Alto, USA  
Hiroshima City University, Japan  
Jozef Stefan Institute, Slovenia  
Hiroshima University, Japan  
Tokyo Institute of Technology, Japan  
University of Waikato, New Zealand  
University of Economics, Czech Republic  
NTT Communication Science Laboratories, Japan  
Université Paris-Sud, France  
National University of Singapore  
Kyushu University, Japan  
Monash University, Australia  
University of Ljubljana, Slovenia  
University of Porto, Portugal  
Kobe University, Japan  
Osaka University, Japan  
Johannes Kepler University, Austria  
Kyoto University, Japan  
Rensselaer Polytechnic Institute, USA  
Czech Technical University in Prague,  
Czech Republic  
Université Paris 13, France

## Local Arrangements

Akira Ishino                      Tohoku University, Japan

## Subreferees

Vineet Chaoji	Saori Kawasaki	Saeed Salem
Jure Ferlež	Attila Kertesz-Farkas	Yacine Sam
Blaz Fortuna	Richard Kirkby	Kazuhiro Seki
Gemma Garriga	Jussi Kujala	Ivica Slavkov
Miha Grcar	Thanh-Phuong Nguyen	Jan-Nikolas Sulzmann
Hakim Hacid	Yoshiaki Okubo	Swee Chuan Tan
Corneliu Henegar	Benjarath Phoophakdee	Shyh Wei Teng
Sachio Hirokawa	Pedro Rodrigues	Yoshiaki Yasumura
Frederik Janssen	Ulrich Rückert	Monika Zakova

# Table of Contents

## Invited Papers

Challenge for Info-plosion .....	1
<i>Masaru Kitsuregawa</i>	
Machine Learning in Ecosystem Informatics .....	9
<i>Thomas G. Dietterich</i>	
Simple Algorithmic Principles of Discovery, Subjective Beauty, Selective Attention, Curiosity & Creativity .....	26
<i>Jürgen Schmidhuber</i>	
A Theory of Similarity Functions for Learning and Clustering .....	39
<i>Avrim Blum</i>	
A Hilbert Space Embedding for Distributions .....	40
<i>Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf</i>	

## Long Papers

Time and Space Efficient Discovery of Maximal Geometric Graphs .....	42
<i>Hiroki Arimura, Takeaki Uno, and Shinichi Shimozono</i>	
Iterative Reordering of Rules for Building Ensembles Without Relearning .....	56
<i>Paulo J. Azevedo and Alípio M. Jorge</i>	
On Approximating Minimum Infrequent and Maximum Frequent Sets .....	68
<i>Mario Boley</i>	
A Partially Dynamic Clustering Algorithm for Data Insertion and Removal .....	78
<i>Haytham Elghazel, Hamamache Kheddouci, Véronique Deslandres, and Alain Dussauchoy</i>	
Positivism Against Constructivism: A Network Game to Learn Epistemology .....	91
<i>Hélène Hagège, Christopher Dartnell, and Jean Sallantin</i>	
Learning Locally Weighted C4.4 for Class Probability Estimation .....	104
<i>Liangxiao Jiang, Harry Zhang, Dianhong Wang, and Zhihua Cai</i>	



User Preference Modeling from Positive Contents for Personalized Recommendation .....	116
<i>Heung-Nam Kim, Inay Ha, Jin-Guk Jung, and Geun-Sik Jo</i>	
Reducing Trials by Thinning-Out in Skill Discovery .....	127
<i>Hayato Kobayashi, Kohei Hatano, Akira Ishino, and Ayumi Shinohara</i>	
A Theoretical Study on Variable Ordering of Zero-Suppressed BDDs for Representing Frequent Itemsets .....	139
<i>Shin-ichi Minato</i>	
Fast NML Computation for Naive Bayes Models .....	151
<i>Tommi Mononen and Petri Myllymäki</i>	
Unsupervised Spam Detection Based on String Alieness Measures .....	161
<i>Kazuyuki Narisawa, Hideo Bannai, Kohei Hatano, and Masayuki Takeda</i>	
A Consequence Finding Approach for Full Clausal Abduction.....	173
<i>Oliver Ray and Katsumi Inoue</i>	
Literature-Based Discovery by an Enhanced Information Retrieval Model.....	185
<i>Kazuhiro Seki and Javed Mostafa</i>	
Discovering Mentorship Information from Author Collaboration Networks .....	197
<i>V. Suresh, Narayanan Raghupathy, B. Shekar, and C.E. Veni Madhavan</i>	
Active Contours as Knowledge Discovery Methods .....	209
<i>Arkadiusz Tomczyk, Piotr S. Szczepaniak, and Michal Pryczek</i>	
An Efficient Polynomial Delay Algorithm for Pseudo Frequent Itemset Mining .....	219
<i>Takeaki Uno and Hiroki Arimura</i>	
Discovering Implicit Feedbacks from Search Engine Log Files .....	231
<i>Ashok Veilumuthu and Parthasarathy Ramachandran</i>	

## Regular Papers

Pharmacophore Knowledge Refinement Method in the Chemical Structure Space .....	243
<i>Satoshi Fujishima, Yoshimasa Takahashi, and Takashi Okada</i>	
An Attempt to Rebuild C. Bernard's Scientific Steps .....	248
<i>Jean-Gabriel Ganascia and Bassel Habib</i>	

Semantic Annotation of Data Tables Using a Domain Ontology . . . . .	253
<i>Gaëlle Hignette, Patrice Buche, Juliette Dibia-Barthélemy, and Olivier Haemmerlé</i>	
Model Selection and Estimation Via Subjective User Preferences . . . . .	259
<i>Jaakko Hollmén</i>	
Detecting Concept Drift Using Statistical Testing . . . . .	264
<i>Kyosuke Nishida and Koichiro Yamauchi</i>	
Towards Future Technology Projection: A Method for Extracting Capability Phrases from Documents . . . . .	270
<i>Risa Nishiyama, Hironori Takeuchi, and Hideo Watanabe</i>	
Efficient Incremental Mining of Top-K Frequent Closed Itemsets . . . . .	275
<i>Andrea Pietracaprina and Fabio Vandin</i>	
An Intentional Kernel Function for RNA Classification . . . . .	281
<i>Hiroshi Sankoh, Koichiro Doi, and Akihiro Yamamoto</i>	
Mining Subtrees with Frequent Occurrence of Similar Subtrees . . . . .	286
<i>Hisashi Tosaka, Atsuyoshi Nakamura, and Mineichi Kudo</i>	
Semantic Based Real-Time Clustering for PubMed Literatures . . . . .	291
<i>Ruey-Ling Yeh, Ching Liu, Ben-Chang Shia, I-Jen Chiang, Wen-Wen Yang, and Hsiang-Chun Tsai</i>	
<b>Author Index</b> . . . . .	297