

Lecture Notes in Artificial Intelligence 2843

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Gunter Grieser Yuzuru Tanaka
Akihiro Yamamoto (Eds.)

Discovery Science

6th International Conference, DS 2003
Sapporo, Japan, October 17-19, 2003
Proceedings



Springer

Series Editors

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

Volume Editors

Gunter Grieser
Technical University Darmstadt
Alexanderstr. 10, 64283 Darmstadt, Germany
E-mail: grieser@informatik.tu-darmstadt.de

Yuzuru Tanaka
Akihiro Yamamoto
Hokkaido University
MemeMedia Laboratory
N-13, W-8, Sapporo, 060-8628, Japan
E-mail: {tanaka;yamamoto}@meme.hokudai.ac.jp

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

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

ISSN 0302-9743

ISBN 3-540-20293-5 Springer-Verlag 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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2003
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin GmbH
Printed on acid-free paper SPIN: 10964132 06/3142 5 4 3 2 1 0

Preface

This volume contains the papers presented at the 6th International Conference on Discovery Science (DS 2003) held at Hokkaido University, Sapporo, Japan, during 17–19 October 2003.

The main objective of the discovery science (DS) conference series is to provide an open forum for intensive discussions and the exchange of new information among researchers working in the area of discovery science. It has become a good custom over the years that the DS conference is held in parallel 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 the good tradition, DS 2003 was co-located with the 14th ALT conference (ALT 2003). The proceedings of ALT 2003 were published as a twin volume 2842 of the LNAI series.

The DS conference series has been supervised by the international steering committee chaired by Hiroshi Motoda (Osaka University, Japan). The other members are Alberto Apostolico (Univ. of Padova, Italy and Purdue University, USA), Setsuo Arikawa (Kyushu University, Japan), Achim Hoffmann (UNSW, Australia), Klaus P. Jantke (DFKI, Germany), Masahiko Sato (Kyoto University, Japan), Ayumi Shinohara (Kyushu University, Japan), Carl H. Smith (University of Maryland, USA), and Thomas Zeugmann (University of Lübeck, Germany).

We received 80 submissions, out of which 18 long and 29 short papers were selected by the program committee, based on clarity, significance, and originality, as well as on relevance to the field of discovery science. The DS 2003 conference had two paper categories, long and short papers. Long papers were presented as 25-minute talks, while short papers were presented as 5-minute talks accompanied by poster presentations. Due to the limited time of the conference, some long submissions could only be accepted as short papers. Some authors of those papers decided not to submit final versions or to just publish abstracts of their papers.

This volume consists of three parts. The first part contains the invited talks of ALT 2003 and DS 2003. These talks were given by Thomas Eiter (Technische Universität Wien, Austria), Genshiro Kitagawa (The Institute of Statistical Mathematics, Japan), Akihiko Takano (National Institute of Informatics, Japan), Naftali Tishby (The Hebrew University, Israel), and Thomas Zeugmann (University of Lübeck, Germany). Because the invited talks were shared by the DS 2003 and ALT 2003 conferences, this volume contains the full versions of Thomas Eiter's, Genshiro Kitagawa's, and Akihiko Takano's talks, as well as abstracts of the talks by the others. The second part of this volume contains the accepted long papers, and the third part contains the accepted short papers.

We would like to express our gratitude to our program committee members and their subreferees who did great jobs in reviewing and evaluating the submissions, and who made the final decisions through intensive discussions to ensure the high quality of the conference. Furthermore, we thank everyone who led this conference to a great success: the authors for submitting papers, the invited speakers for accepting our invitations and giving stimulating talks, the steering committee and the sponsors for their support, the ALT Chairpersons Klaus P. Jantke, Ricard Gavalda, and Eiji Takimoto for their collaboration, and, last but not least, Makoto Haraguchi and Yoshiaki Okubo (both Hokkaido University, Japan) for their local arrangement of the twin conferences.

October 2003

Gunter Grieser
Yuzuru Tanaka
Akihiro Yamamoto

Organization

Conference Chair

Yuzuru Tanaka Hokkaido University, Japan

Program Committee

Gunter Grieser (Co-chair)	Technical University, Darmstadt, Germany
Akihiro Yamamoto (Co-chair)	Hokkaido University, Japan
Simon Colton	Imperial College London, UK
Vincent Corruble	Université P. et M. Curie Paris, France
Johannes Fürnkranz	Research Institute for AI, Austria
Achim Hoffmann	University of New South Wales, Australia
Naresh Iyer	GE Global Research Center, USA
John R. Josephson	Ohio State University, USA
Eamonn Keogh	University of California, USA
Mineichi Kudo	Hokkaido University, Japan
Nicolas Lachiche	University of Strasbourg, France
Steffen Lange	DFKI GmbH, Germany
Lorenzo Magnani	University of Pavia, Italy
Michael May	Fraunhofer AIS, Germany
Hiroshi Motoda	Osaka University, Japan
Nancy J. Nersessian	Georgia Institute of Technology, USA
Vladimir Pericliev	Academy of Sciences, Sofia, Bulgaria
Jan Rauch	University of Economics, Prague, Czech Republic
Henk W. de Regt	Free University Amsterdam, The Netherlands
Ken Satoh	National Institute of Informatics, Japan
Tobias Scheffer	Humboldt University, Berlin, Germany
Einoshin Suzuki	Yokohama National University, Japan
Masayuki Takeda	Kyushu University, Japan
Ljupčo Todorovski	Jožef Stefan Institute, Ljubljana, Slovenia
Gerhard Widmer	University of Vienna, Austria

Local Arrangements

Makoto Haraguchi (Chair) Hokkaido University, Japan
Yoshiaki Okubo Hokkaido University, Japan

Subreferees

Hideo Bannai
Rens Bod
Marco Chiarandini
Nigel Collier
Pascal Divoux
Thomas Gärtner
Peter Grigoriev
Makoto Haraguchi
Katsutoshi Hirayama
Akira Ishino
Takayuki Ito
Ai Kawazoe
Nikolay Kirov
Edda Leopold

Tsuyoshi Murata
Atsuyoshi Nakamura
Luis F. Paquete
Lourdes Peña Castillo
Johann Petrak
Detlef Prescher
Ehud Reither
Alexandr Savinov
Tommaso Schiavinotto
Alexander K. Seewald
Esko Ukkonen
Serhiy Yevtushenko
Kenichi Yoshida
Sandra Zilles

Sponsors

Japanese Ministry of Education, Culture, Sports, Science and Technology
(MEXT)
The Suginome Memorial Foundation, Japan

Table of Contents

Invited Talks

Abduction and the Dualization Problem	1
<i>Thomas Eiter, Kazuhisa Makino</i>	
Signal Extraction and Knowledge Discovery Based on Statistical Modeling	21
<i>Genshiro Kitagawa</i>	
Association Computation for Information Access	33
<i>Akihiko Takano</i>	
Efficient Data Representations That Preserve Information	45
<i>Naftali Tishby</i>	
Can Learning in the Limit Be Done Efficiently?	46
<i>Thomas Zeugmann</i>	

Long Papers

Discovering Frequent Substructures in Large Unordered Trees	47
<i>Tatsuya Asai, Hiroki Arimura, Takeaki Uno, Shin-ichi Nakano</i>	
Discovering Rich Navigation Patterns on a Web Site	62
<i>Karine Chevalier, Cécile Bothorel, Vincent Corruble</i>	
Mining Frequent Itemsets with Category-Based Constraints	76
<i>Tien Dung Do, Siu Cheung Hui, Alvis Fong</i>	
Modelling Soil Radon Concentration for Earthquake Prediction	87
<i>Sašo Džeroski, Ljupčo Todorovski, Boris Zmazek, Janja Vaupotič, Ivan Kopal</i>	
Dialectical Evidence Assembly for Discovery	100
<i>Alistair Fletcher, John Davis</i>	
Performance Analysis of a Greedy Algorithm for Inferring Boolean Functions	114
<i>Daiji Fukagawa, Tatsuya Akutsu</i>	
Performance Evaluation of Decision Tree Graph-Based Induction	128
<i>Warodom Geamsakul, Takashi Matsuda, Tetsuya Yoshida, Hiroshi Motoda, Takashi Washio</i>	

Discovering Ecosystem Models from Time-Series Data	141
<i>Dileep George, Kazumi Saito, Pat Langley, Stephen Bay, Kevin R. Arrigo</i>	
An Optimal Strategy for Extracting Probabilistic Rules by Combining Rough Sets and Genetic Algorithm	153
<i>Xiaoshu Hang, Honghua Dai</i>	
Extraction of Coverings as Monotone DNF Formulas	166
<i>Kouichi Hirata, Ryosuke Nagazumi, Masateru Harao</i>	
What Kinds and Amounts of Causal Knowledge Can Be Acquired from Text by Using Connective Markers as Clues?	180
<i>Takashi Inui, Kentaro Inui, Yuji Matsumoto</i>	
Clustering Orders	194
<i>Toshihiro Kamishima, Jun Fujiki</i>	
Business Application for Sales Transaction Data by Using Genome Analysis Technology	208
<i>Naoki Katoh, Katsutoshi Yada, Yukinobu Hamuro</i>	
Improving Efficiency of Frequent Query Discovery by Eliminating Non-relevant Candidates	220
<i>Jérôme Maloberti, Einoshin Suzuki</i>	
Chaining Patterns	233
<i>Taneli Mielikäinen</i>	
An Algorithm for Discovery of New Families of Optimal Regular Networks	245
<i>Oleg Monakhov, Emilia Monakhova</i>	
Enumerating Maximal Frequent Sets Using Irredundant Dualization	256
<i>Ken Satoh, Takeaki Uno</i>	
Discovering Exceptional Information from Customer Inquiry by Association Rule Miner	269
<i>Keiko Shimazu, Atsuhito Momma, Koichi Furukawa</i>	
Short Papers	
Automatic Classification for the Identification of Relationships in a Meta-data Repository	283
<i>Gerd Beuster, Ulrich Furbach, Margret Gross-Hardt, Bernd Thomas</i>	
Effects of Unreliable Group Profiling by Means of Data Mining	291
<i>Bart Custers</i>	

Using Constraints in Discovering Dynamics	297
<i>Sašo Džeroski, Ljupčo Todorovski, Peter Ljubič</i>	
SA-Optimized Multiple View Smooth Polyhedron Representation NN	306
<i>Mohamad Ivan Fanany, Itsuo Kumazawa</i>	
Elements of an Agile Discovery Environment	311
<i>Peter A. Grigoriev, Serhiy A. Yevtushenko</i>	
Discovery of User Preference in Personalized Design Recommender System through Combining Collaborative Filtering and Content Based Filtering	320
<i>Kyung-Yong Jung, Jason J. Jung, Jung-Hyun Lee</i>	
Discovery of Relationships between Interests from Bulletin Board System by Dissimilarity Reconstruction	328
<i>Kou Zhongbao, Ban Tao, Zhang Changshui</i>	
A Genetic Algorithm for Inferring Pseudoknotted RNA Structures from Sequence Data	336
<i>Dongkyu Lee, Kyungsook Han</i>	
Prediction of Molecular Bioactivity for Drug Design Using a Decision Tree Algorithm	344
<i>Sanghoon Lee, Jihoon Yang, Kyung-whan Oh</i>	
Mining RNA Structure Elements from the Structure Data of Protein-RNA Complexes	352
<i>Daeho Lim, Kyungsook Han</i>	
Discovery of Cellular Automata Rules Using Cases	360
<i>Ken-ichi Maeda, Chiaki Sakama</i>	
Discovery of Web Communities from Positive and Negative Examples	369
<i>Tsuyoshi Murata</i>	
Association Rules and Dempster-Shafer Theory of Evidence	377
<i>Tetsuya Murai, Yasuo Kudo, Yoshiharu Sato</i>	
Subgroup Discovery among Personal Homepages	385
<i>Toyohisa Nakada, Susumu Kunifuji</i>	
Collaborative Filtering Using Projective Restoration Operators	393
<i>Atsuyoshi Nakamura, Mineichi Kudo, Akira Tanaka, Kazuhiko Tanabe</i>	
Discovering Homographs Using N-Partite Graph Clustering	402
<i>Hidekazu Nakawatase, Akiko Aizawa</i>	

Discovery of Trends and States in Irregular Medical Temporal Data	410
<i>Trong Dung Nguyen, Saori Kawasaki, Tu Bao Ho</i>	
Creating Abstract Concepts for Classification by Finding Top- <i>N</i> Maximal Weighted Cliques	418
<i>Yoshiaki Okubo, Makoto Haraguchi</i>	
Content-Based Scene Change Detection of Video Sequence Using Hierarchical Hidden Markov Model	426
<i>Jong-Hyun Park, Soon-Young Park, Seong-Jun Kang, Wan-Hyun Cho</i>	
An Appraisal of UNIVAUTO – The First Discovery Program to Generate a Scientific Article	434
<i>Vladimir Perichiev</i>	
Scilog: A Language for Scientific Processes and Scales	442
<i>Joseph Phillips</i>	
Mining Multiple Clustering Data for Knowledge Discovery	452
<i>Thanh Tho Quan, Siu Cheung Hui, Alvis Fong</i>	
Bacterium Lingualis – The Web-Based Commonsensical Knowledge Discovery Method	460
<i>Rafal Rzepka, Kenji Araki, Koji Tochinai</i>	
Inducing Biological Models from Temporal Gene Expression Data	468
<i>Kazumi Saito, Dileep George, Stephen Bay, Jeff Shrager</i>	
Knowledge Discovery on Chemical Reactivity from Experimental Reaction Information	470
<i>Hiroko Satoh, Tadashi Nakata</i>	
A Method of Extracting Related Words Using Standardized Mutual Information	478
<i>Tomohiko Sugimachi, Akira Ishino, Masayuki Takeda, Fumihiko Matsuo</i>	
Discovering Most Classificatory Patterns for Very Expressive Pattern Classes	486
<i>Masayuki Takeda, Shunsuke Inenaga, Hideo Bannai, Ayumi Shinohara, Setsuo Arikawa</i>	
Mining Interesting Patterns Using Estimated Frequencies from Subpatterns and Superpatterns	494
<i>Yukiko Yoshida, Yuiko Ohta, Ken'ichi Kobayashi, Nobuhiro Yugami</i>	
Author Index	503