Preface

This year’s International Conference on Discovery Science (DS) was the 16th event in this series. Like in previous years, the conference was co-located with the International Conference on Algorithmic Learning Theory (ALT), which is already in its 24th year. Starting in 2001, ALT/DS is one of the longest-running series of co-located events in computer science. The unique combination of recent advances in the development and analysis of methods for automatic scientific knowledge discovery, machine learning, intelligent data analysis, and their application to knowledge discovery on the one hand, and theoretical and algorithmic advances in machine learning on the other hand makes every instance of this joint event unique and attractive.

This volume contains the papers presented at the 16th International Conference on Discovery Science, while the papers of the 24th International Conference on Algorithmic Learning Theory are published in a companion volume edited by Sanjay Jain, Rémi Munos, Frank Stephan, and Thomas Zeugmann (Springer LNCS Vol. 8139). We had the pleasure of selecting contributions from 52 submissions by 142 authors from 23 countries. Each submission was reviewed by three Program Committee members. The program chairs eventually decided to accept 23 papers, yielding an acceptance rate of slightly less than 45%.

The program also included 3 invited talks and 2 tutorials. In the joint DS/ALT invited talk, Nir Ailon gave a presentation about “Learning and Optimizing with Preferences.” The DS invited talk by Hannu Toivonen was on “Creative Computers and Data Mining.” Finally, DS participants also had the opportunity to attend the ALT invited talk on “Efficient Algorithms for Combinatorial Online Prediction”, which was given by Eiji Takimoto. The two tutorial speakers were Krzysztof Dembczyński (“Multi-Target Prediction”) and Nader H. Bshouty (“Exact Learning from Membership Queries: Some Techniques, Results and New Directions”).

This year, both conferences were held in Singapore, organized by the School of Computing, National University of Singapore (NUS). We are very grateful to the School of Computing at NUS for sponsoring the conferences and providing administrative support. In particular, we thank the local arrangement chair, Lee Wee Sun, and his team, Mark Bartholomeusz and Kee Yong Ngee, as well as all of the other administrative staff at the School of Computing, NUS, for their efforts in organizing the two conferences. We would like to thank the Office of Naval Research Global for the generous financial support provided under ONRG GRANT N62909-13-1-C208.

We would also like to thank 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 are grateful to Frank Stephan and Sanjay Jain for their timely answers to many
questions and for ensuring a smooth coordination with ALT, Thomas Zeugmann for his help with the proceedings, Robin Senge for putting up and maintaining our website, and Andrei Voronkov for making EasyChair freely available. Finally, special thanks go to the Discovery Science Steering Committee, in particular to its past and current chairs, Einoshin Suzuki and Akihiro Yamamoto, for entrusting us with the organization of the scientific program of this prestigious conference.

July 2013

Johannes Fürnkranz
Eyke Hüllermeier
Tomoyuki Higuchi
Organization

Conference Chair
Tomoyuki Higuchi
The Institute of Statistical Mathematics, Tokyo

Program Chairs
Johannes Fürnkranz
TU Darmstadt, Germany
Eyke Hüllermeier
University of Marburg, Germany

Local Arrangements Chair
Wee Sun Lee

Local Organization Team
Frank Stephan
Sanjay Jain
Mark Bartholomeusz
Yong Ngee Kee
Rachel Goh
Noraiszah Bte Hamzah

Program Committee
Fabio Aiolfi
Università di Padova, Italy
Hiroki Arimura
Hokkaido University, Japan
Hideo Bannai
Kyushu University, Japan
Michael Berthold
University of Konstanz, Germany
Albert Bifet
University of Waikato, New Zealand
Henrik Blockeel
Katholieke Universiteit Leuven, Belgium
Henrik Boström
Stockholm University, Sweden
Jean-Francois Boulicaut
INSA Lyon, France
Ulf Brefeld
TU Darmstadt, Germany
Robert Busa-Fekete
University of Marburg, Germany
Weiwei Cheng
University of Marburg, Germany
Bruno Cremilleux
University of Caen, France
Luc De Raedt
Katholieke Universiteit Leuven, Belgium
Juan José Del Coz
University of Oviedo at Gijón, Spain
Krzysztof Dembczyński
Poznan University of Technology, Poland
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sašo Džeroski</td>
<td>Jozef Stefan Institute, Slovenia</td>
</tr>
<tr>
<td>Tapio Elomaa</td>
<td>Tampere University of Technology, Finland</td>
</tr>
<tr>
<td>Ad Feelders</td>
<td>Universiteit Utrecht, The Netherlands</td>
</tr>
<tr>
<td>Peter Flach</td>
<td>University of Bristol, UK</td>
</tr>
<tr>
<td>Joao Gama</td>
<td>University Porto, Portugal</td>
</tr>
<tr>
<td>Mohand-Said Hacid</td>
<td>Université Claude Bernard Lyon, France</td>
</tr>
<tr>
<td>Howard Hamilton</td>
<td>University of Regina, Canada</td>
</tr>
<tr>
<td>Makoto Haraguchi</td>
<td>Hokkaido University, Japan</td>
</tr>
<tr>
<td>Kouichi Hirata</td>
<td>Kyushu Institute of Technology, Japan</td>
</tr>
<tr>
<td>Jaakko Hollmén</td>
<td>Aalto University School of Science, Finland</td>
</tr>
<tr>
<td>Geoffrey Holmes</td>
<td>University of Waikato, New Zealand</td>
</tr>
<tr>
<td>Tamas Horvath</td>
<td>University of Bonn and Fraunhofer IAIS, Germany</td>
</tr>
<tr>
<td>Alípio M. Jorge</td>
<td>University of Porto, Portugal</td>
</tr>
<tr>
<td>Hisashi Kashima</td>
<td>The University of Tokyo, Japan</td>
</tr>
<tr>
<td>Kristian Kersting</td>
<td>Fraunhofer IAIS and University of Bonn, Germany</td>
</tr>
<tr>
<td>Ross King</td>
<td>University of Manchester, UK</td>
</tr>
<tr>
<td>Arno Knobbe</td>
<td>Universiteit Leiden, The Netherlands</td>
</tr>
<tr>
<td>Joost Kok</td>
<td>LIACS, Leiden University, The Netherlands</td>
</tr>
<tr>
<td>Stefan Kramer</td>
<td>Johannes Gutenberg University Mainz, Germany</td>
</tr>
<tr>
<td>Nada Lavrač</td>
<td>Jozef Stefan Institute, Slovenia</td>
</tr>
<tr>
<td>Philippe Lenca</td>
<td>Telecom Bretagne, France</td>
</tr>
<tr>
<td>Ee-Peng Lim</td>
<td>Singapore Management University, Singapore</td>
</tr>
<tr>
<td>Eneldo Loza Mencia</td>
<td>TU Darmstadt, Germany</td>
</tr>
<tr>
<td>Donato Malerba</td>
<td>Università di Bari, Italy</td>
</tr>
<tr>
<td>Taneli Miilikäinen</td>
<td>Nokia Research Center, USA</td>
</tr>
<tr>
<td>Richard Nock</td>
<td>CEREGMIA-UAG, France</td>
</tr>
<tr>
<td>Panagiotis Papapetrou</td>
<td>University of London, UK</td>
</tr>
<tr>
<td>Mykola Pechenizkiy</td>
<td>Eindhoven University of Technology, The Netherlands</td>
</tr>
<tr>
<td>Jean-Marc Petit</td>
<td>University of Lyon, INSA Lyon, France</td>
</tr>
<tr>
<td>Bernhard Pfahringer</td>
<td>University of Waikato, New Zealand</td>
</tr>
<tr>
<td>Marc Plantevit</td>
<td>University of Lyon, France</td>
</tr>
<tr>
<td>Enric Plaza</td>
<td>IIIA-CSIC, Spain</td>
</tr>
<tr>
<td>Vítor Santos Costa</td>
<td>Universidade do Porto, Portugal</td>
</tr>
<tr>
<td>Mika Sulkava</td>
<td>MTT Agrifood Research Finland, Finland</td>
</tr>
<tr>
<td>Einoshin Suzuki</td>
<td>Kyushu University, Japan</td>
</tr>
<tr>
<td>Maguelonne Teisseire</td>
<td>Cemagref - UMR Tetis, France</td>
</tr>
<tr>
<td>Grigorios Tsoumakas</td>
<td>Aristotle University, Greece</td>
</tr>
<tr>
<td>Xizhao Wang</td>
<td>Hebei University, China</td>
</tr>
<tr>
<td>Stefan Wrobel</td>
<td>Fraunhofer IAIS and University of Bonn, Germany</td>
</tr>
<tr>
<td>Yang Yu</td>
<td>Nanjing University, China</td>
</tr>
<tr>
<td>Min-Ling Zhang</td>
<td>Southeast University, China</td>
</tr>
<tr>
<td>Zhi-Hua Zhou</td>
<td>Nanjing University, China</td>
</tr>
<tr>
<td>Albrecht Zimmermann</td>
<td>Katholieke Universiteit Leuven, Belgium</td>
</tr>
</tbody>
</table>
Additional Reviewers

Alatrista Salas, Hugo  Moreira-Matias, Luis
Appice, Annalisa  Otaki, Keisuke
Bothorel, Cecile  Panov, Pance
Cerf, Loïc  Phan, Nhat Hai
Desmier, Elise  Rabatel, Julien
Hoang, Lam Thanh  Ribeiro, Rita
Huang, Sheng-Jun  Sakamoto, Hiroshi
Kaytoue, Mehdi  Soulet, Arnaud
Kocev, Dragi  Tromp, Erik
Kuboyama, Tetsuji  Zenko, Bernard
Kuzmanovski, Vladimir
Invited Talks
(Abstracts)
Learning and Optimizing with Preferences*

Nir Ailon

Department of Computer Science
Technion Israel Institute of Technology
Haifa, Israel
nailon@cs.technion.ac.il

Abstract. Preferences and choices are a central source of information generated by humans. They have been studied for centuries in the context of social choice theory, econometric theory, statistics and psychology. At least two Nobel prizes in economics have been awarded for work reasoning about human preferences and choices.

In the last two decades computer scientists have studied preference data, which became available in unprecedented quantities: Each time we click or tap on a search result, a sponsored ad or a product recommendation, we express preference of one alternative from a small set of alternatives. Additionally, many crowdsourcing systems explicitly ask (paid?) experts to solicit preferences or even full rankings of alternative sets.

What are the advantages of preferences compared to other forms of information, and what challenges do they give rise to? I will present important problems and survey results.

Efficient Algorithms for Combinatorial Online Prediction*

Eiji Takimoto

Department of Informatics
Kyushu University
Japan
eiji@inf.kyushu-u.ac.jp

We study online linear optimization problems over combinatorial concept classes $C \subseteq \mathbb{R}^n$ that are defined in some combinatorial ways. Examples of such classes are $s$-$t$ paths in a given graph, spanning trees of a given graph, permutations over a given set, truth assignments for a given CNF formula, set covers of a given subset family, and so on. Typically, those concept classes are finite but contain exponentially many concepts. The problem for a concept class $C$ is described as follows: At each trial $t$, the algorithm chooses a concept $c_t \in C$, the adversary returns a loss vector $\ell_t \in [0,1]^n$, and the algorithm incurs a loss given by $c_t \cdot \ell_t$. The goal of the algorithm is to make the cumulative loss not much larger than that of the best concept in $C$.

One of the major approaches to the problem is to apply Follow the Regularized Leader (FTRL) framework, in which two external procedures projection and decomposition are assumed to be implemented. In other words, for each concept class $C$, we need to design algorithms for the two procedures. In this talk, we give a projection and decomposition algorithms that work efficiently and uniformly for a wide class of concept classes. More precisely, if the convex hull of $C$ is a submodular base polyhedron specified by a submodular function $f$, then the two procedures are computed in polynomial time, assuming that $f$ can be computed in polynomial time.

Another approach is to use an offline algorithm as an oracle to construct an online algorithm. Here, the offline algorithm solves the corresponding offline optimization problem. Follow the perturbed leader (FPL) and the Online Frank-Wolfe (OFW) are of this type. In this talk, we consider a harder but typical case where the offline optimization problem for $C$ is NP-hard, for which none of the FTRL, FPL and OFW work. The FTRL has been generalized so that it works when an offline approximation algorithm is available. However, it is not efficient enough. In this talk, we give a more efficient online algorithm using an offline approximation algorithm which has a guarantee of a certain integrity gap.

* An extended version of this paper can be found in Sanjay Jain, Rémi Munos, Frank Stephan, and Thomas Zeugmann, Proceedings of the 24th International Conference on Algorithmic Learning Theory (ALT-13), Lecture Notes in Computer Science Vol. 8139, Springer-Verlag, 2013.
Abstract. In the field of computational creativity, researchers aim to give computers creative skills, such as those needed in writing poetry or composing music. Obviously, an agent needs to know the field in which it operates. This is where data mining has great potential: making creative agents adaptive to various fields and genres by automatic discovery of relevant information from existing creative artifacts. We give several examples of how verbal creativity can benefit from data mining of existing text corpora.

On the other hand, computational creativity tools allow a whole new approach to data analysis. In this “Affective Data Analysis”, the goal is to turn data into a subjective, esthetic experience by automatic or semiautomatic creation of a novel artifact using the user’s data as inspiration. This is in strong contrast with traditional data analysis methods that emphasize cold facts instead of warm feelings. We illustrate this idea with musicalization of sleep measurements and chat discussions.
# Table of Contents

Mixture Models from Multiresolution 0-1 Data ............................... 1  
*Prem Raj Adhikari and Jaakko Hollmén*

Model Tree Ensembles for Modeling Dynamic Systems ...................... 17  
*Darko Aleksovski, Juš Kocijan, and Sašo Džeroski*

Fast and Scalable Image Retrieval Using Predictive Clustering Trees .... 33  
*Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, and Sašo Džeroski*

Avoiding Anomalies in Data Stream Learning ............................... 49  
*João Gama, Petr Kosina, and Ezilda Almeida*

Generalizing from Example Clusters ........................................ 64  
*Pan Hu, Celine Vens, Bart Verstrynge, and Hendrik Blockeel*

Clustering Based Active Learning for Evolving Data Streams .......... 79  
*Dino Ienco, Albert Bifet, Indrė Žliobaitė, and Bernhard Pfahringer*

Robust Crowd Labeling Using Little Expertise ........................... 94  
*Faiza Khan Khattak and Ansaf Salleb-Aouissi*

A New Approach to String Pattern Mining with Approximate Match ... 110  
*Tetsushi Matsui, Takeaki Uno, Juzoh Umemori, and Tsuyoshi Koide*

OntoDM-KDD: Ontology for Representing the Knowledge Discovery Process ................................................................. 126  
*Panče Panov, Larisa Soldatova, and Sašo Džeroski*

A Wordification Approach to Relational Data Mining ..................... 141  
*Matic Perovšek, Anže Vavpetič, Bojan Cestnik, and Nada Lavrač*

Multi-interval Discretization of Continuous Attributes for Label Ranking ................................................................. 155  
*Cláudio Rebelo de Sá, Carlos Soares, Arno Knobbe,  
Paulo Azevedo, and Alípio Mário Jorge*

Identifying Super-Mediators of Information Diffusion in Social Networks ................................................................. 170  
*Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda*

SM2D: A Modular Knowledge Discovery Approach Applied to Hydrological Forecasting ............................................... 185  
*Wilfried Segretier and Martine Collard*
A Dynamic Programming Algorithm for Learning Chain Event Graphs .......................................................... 201
  Tomi Silander and Tze-Yun Leong

Mining Interesting Patterns in Multi-relational Data with N-ary Relationships .................................................... 217
  Eirini Spyropoulou, Tijl De Bie, and Mario Boley

Learning Hierarchical Multi-label Classification Trees from Network Data ........................................................... 233
  Daniela Stojanova, Michelangelo Ceci, Donato Malerba, and Sašo Džeroski

A Density-Based Backward Approach to Isolate Rare Events in Large-Scale Applications ........................................ 249
  Énikő Székely, Pascal Poncelet, Florent Masseglia, Maguelonne Teisseire, and Renaud Cezar

Inductive Process Modeling of Rab5-Rab7 Conversion in Endocytosis ... 265
  Jovan Tanevski, Ljupčo Todorovski, Yannis Kalaidzidis, and Sašo Džeroski

Fast Compression of Large-Scale Hypergraphs for Solving Combinatorial Problems .................................................... 281
  Takahisa Toda

Semantic Data Mining of Financial News Articles .......................... 294
  Anže Vavpetič, Petra Kralj Novak, Miha Grčar, Igor Mozetič, and Nada Lavrač

Polynomial Delay and Space Discovery of Connected and Acyclic Sub-hypergraphs in a Hypergraph .......................... 308
  Kunihiro Wasa, Takeaki Uno, Kouichi Hirata, and Hiroki Arimura

Hyperlink Prediction in Hypernetworks Using Latent Social Features ... 324
  Ye Xu, Dan Rockmore, and Adam M. Kleinbaum

Extracting Opinionated (Sub)Features from a Stream of Product Reviews ................................................................. 340
  Max Zimmermann, Eirini Ntoutsi, and Myra Spiliopoulou

Author Index ........................................................................................................... 357