

Lecture Notes in Artificial Intelligence 7523

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

Peter A. Flach Tijn De Bie
Nello Cristianini (Eds.)

Machine Learning and Knowledge Discovery in Databases

European Conference, ECML PKDD 2012
Bristol, UK, September 24-28, 2012
Proceedings, Part I



Springer

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

Peter A. Flach
Tijl De Bie
Nello Cristianini
University of Bristol
Intelligent Systems Laboratory
Merchant Venturers Building
Woodland Road
Bristol BS8 1UB, UK
E-mails:
peter.flach@bristol.ac.uk
tijl.debie@bristol.ac.uk
nello.cristianini@bristol.ac.uk

© Cover illustration by www.zedphoto.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-33459-7 e-ISBN 978-3-642-33460-3
DOI 10.1007/978-3-642-33460-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012946760

CR Subject Classification (1998):
I.2.6, H.2.8, I.5.2, G.2.2, G.3, I.2.4, I.2.7, H.3.4-5, I.2.9, F.2.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2012

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

These proceedings contain the technical papers presented at the 2012 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2012), held in Bristol, UK, during the week of 24–28 September 2012. ECML-PKDD is a highly successful and selective international conference series, which was first organised in its present form in 2001 in Freiburg, Germany, when it joined together the hitherto separate ECML and PKDD conferences. Since then, the two strands of machine learning and data mining have been increasingly integrated in the joint conference, and today it is well-known as the only top-tier international conference that exploits the synergy between these two exciting fields. It is therefore particularly pleasing that the achieved level of synergy is evident from the list of topics in which the research track papers are categorised in these pages: Association Rules and Frequent Patterns; Bayesian Learning and Graphical Models; Classification; Dimensionality Reduction, Feature Selection and Extraction; Distance-Based Methods and Kernels; Ensemble Methods; Graph and Tree Mining; Large-Scale, Distributed and Parallel Mining and Learning; Multi-relational Mining and Learning; Multi-task Learning; Natural Language Processing; Online Learning and Data Streams; Privacy and Security; Rankings and Recommendations; Reinforcement Learning and Planning; Rule Mining and Subgroup Discovery; Semi-supervised and Transductive Learning; Sensor Data; Sequence and String Mining; Social Network Mining; Spatial and Geographical Data Mining; Statistical Methods and Evaluation; Time Series and Temporal Data Mining; and Transfer Learning.

The format of the 2012 conference follows the tried-and-tested format of previous instalments, with workshops and tutorials on Monday and Friday; research papers in parallel tracks on Tuesday, Wednesday and Thursday; and plenary keynote talks on each of the five conference days. The proceedings start with abstracts and bios of our five eminent invited speakers: Pieter Abbeel, Luc De Raedt, Douglas Eck, Daniel Keim and Padhraic Smyth. The bulk of the proceedings is then taken up by 105 research papers. These were carefully selected from 443 submitted papers (acceptance rate 23.7%) on the basis of reviews by 275 Programme Committee members and 36 Area Chairs, assisted by 161 additional reviewers. In acknowledgment of their essential contribution to the success of the conference you will find their names on the following pages.

The final sections of the proceedings are devoted to Demo and Nectar papers. Ten system demonstration papers were selected from 19 submissions to the Demo track by a small committee chaired by Bettina Berendt and Myra Spiliopoulou. The Nectar track is new this year and features significant machine learning and data mining results published or disseminated no earlier than 2010 at different conferences or in journals. One goal of this track is to offer conference attendees the opportunity to learn about results published in other communities but related to machine learning or data mining (or both). Submissions compactly

presenting well-founded results which appeared in a series of publications that advanced a single novel influential idea or vision were also welcomed. A dedicated committee chaired by Gemma Garriga and Thomas Gärtner selected 4 Nectar papers from 14 submissions. Our sincere thanks to everyone involved for these valuable additions to the conference.

Elements of the conference not directly represented in the proceedings include 11 workshops (Mining Ubiquitous and Social Environments; New Frontiers in Mining Complex Patterns; The Silver Lining – Learning from Unexpected Results; Instant Interactive Data Mining; Learning and Discovery in Symbolic Systems Biology; Sentiment Discovery from Affective Data; Active Learning in Real-world Applications; Mining and Exploiting Interpretable Local Patterns; Community Mining and People Recommenders; Collective Learning and Inference on Structured Data: and the Discovery Challenge workshop), as well as 8 tutorials (Understanding and Managing Cascades on Large Graphs; Advanced Topics in Data Stream Mining; Mining Deep Web Repositories; PAC-Bayesian Analysis in Supervised, Unsupervised, and Reinforcement Learning; Random Projections for Machine Learning and Data Mining; Decomposing Binary Matrices; Advanced Topics in Ensemble Learning; and Probabilistic Modeling of Ranking). Many thanks to the workshop organisers and tutorial presenters, as well as the Workshop Chairs Arno Knobbe and Carlos Soares and the Tutorial Chairs Alessandro Moschitti and Siegfried Nijssen for putting together this exciting programme. We would also like to draw attention to the programme of presentations by representatives from industry put together by Industry Track Chairs Cédric Archambeau and David Barber, consisting of a series of talks centred around Big Data as well as a programme of ‘Startup Stories’ sponsored by the PASCAL2 Network of Excellence.

Finally, it is our pleasure to announce the winners of the best paper awards, as selected by a small committee chaired by the Awards Chair Johannes Fürnkranz. The paper ‘Active Evaluation of Ranking Functions based on Graded Relevance’ by Christoph Sawade, Steffen Bickel, Timo von Oertzen, Tobias Scheffer and Niels Landwehr wins the best machine learning paper award sponsored by the *Machine Learning* journal. The paper ‘Socioscope: Spatio-Temporal Signal Recovery from Social Media’ by Jun-Ming Xu, Aniruddha Bhargava, Robert Nowak and Xiaojin Zhu receives the best data mining paper award sponsored by *Data Mining and Knowledge Discovery*. Our congratulations to the authors of both papers and in particular to their student main authors. We also continue the tradition started at ECML-PKDD 2011 in Athens of selecting a most influential paper published at the conference 10 years ago. Following a poll organised by the Awards Chair among selected participants the most influential paper presented at ECML-PKDD 2002 in Helsinki is ‘Mining All Non-derivable Frequent Itemsets’ by Toon Calders and Bart Goethals. We look forward to finding out which paper from these proceedings will be deemed to have been most influential in 2022!

July 2012

Peter Flach
Tijl De Bie
Nello Cristianini

Organisation

ECML-PKDD 2012 was organised by the Departments of Computer Science and Engineering Mathematics of the University of Bristol, UK. We gratefully acknowledge the financial support of our academic sponsors (the University of Bristol (the aforementioned Departments as well as the Faculty of Engineering), the PASCAL2 Network of Excellence and the EternalS FP7 Coordination Action), our business sponsors (Winton Capital Management, Hewlett-Packard Research Labs Bristol, Rapid-I, Cambridge University Press, Google, KNIME, IBM UK and Microsoft Research) and our award sponsors (the journals of *Machine Learning* and *Data Mining and Knowledge Discovery*).

General and Programme Chairs

Peter Flach	University of Bristol, UK
Tijl De Bie	University of Bristol, UK
Nello Cristianini	University of Bristol, UK

Local Organisers

Oliver Ray	University of Bristol, UK
Tilo Burghardt	University of Bristol, UK
Nanlin Jin	University of Bristol, UK
Yizhao Ni	University of Bristol, UK
Simon Price	University of Bristol, UK

Workshop Chairs

Arno Knobbe	Universiteit Leiden, The Netherlands
Carlos Soares	Universidade do Porto, Portugal

Tutorial Chairs

Alessandro Moschitti	University of Trento, Italy
Siegfried Nijssen	Katholieke Universiteit Leuven, Belgium

Demo Track Chairs

Bettina Berendt	Katholieke Universiteit Leuven, Belgium
Myra Spiliopoulou	University of Magdeburg, Germany

Nectar Track Chairs

Gemma C. Garriga	INRIA Lille, France
Thomas Gärtner	University of Bonn & Fraunhofer, Germany

Industry Track Chairs

Cédric Archambeau	Xerox Research Centre Europe, France
David Barber	University College London, UK

Awards Chair

Johannes Fürnkranz	TU Darmstadt, Germany
--------------------	-----------------------

Sponsorship Chairs

Toon Calders	Eindhoven University, The Netherlands
Trevor Martin	University of Bristol, UK

Publicity Chair

Grigorios Tsoumakas	Aristotle University of Thessaloniki, Greece
---------------------	--

Proceedings Chairs

Ilias Flaounas	University of Bristol, UK
Tim Kovacs	University of Bristol, UK

Webmaster

Thomas Lansdall-Welfare	University of Bristol, UK
-------------------------	---------------------------

Discovery Challenge Organisers

Ion Androutsopoulos	Athens University of Economics and Business, Greece
Thierry Artieres	Laboratoire d'Informatique de Paris 6, France
Patrick Gallinari	Laboratoire d'Informatique de Paris 6, France
Eric Gaussier	Laboratoire d'Informatique de Grenoble, France
Aris Kosmopoulos	NCRS "Demokritos" & Athens University of Economics and Business, Greece
George Paliouras	NCRS "Demokritos", Greece
Ioannis Partalas	Laboratoire d'Informatique de Grenoble, France

ECML-PKDD Steering Committee

Fosca Giannotti, chair	Università di Pisa, Italy
Francesco Bonchi	Yahoo! Research Barcelona, Spain
Wray Buntine	NICTA Canberra, Australia
Dimitrios Gunopulos	University of Athens, Greece
Donato Malerba	Università degli Studi di Bari, Italy
Dunja Mladenić	Jožef Stefan Institute, Slovenia
John Shawe-Taylor	University College London, UK
Michèle Sebag	Laboratoire de Recherche en Informatique, France
Michalis Vazirgiannis	Athens University of Economics and Business, Greece

Area Chairs

Annalisa Appice	Università degli Studi di Bari, Italy
Roberto J. Bayardo	Google, USA
Tanya Berger-Wolf	University of Illinois, USA
Hendrik Blockeel	Katholieke Universiteit Leuven, Belgium
Francesco Bonchi	Yahoo! Research Barcelona, Spain
Carla E. Brodley	Tufts University, USA
Carlotta Domeniconi	George Mason University, USA
Tina Eliassi-Rad	Rutgers University, USA
Charles Elkan	University of California San Diego, USA
Tapio Elomaa	Tampere University of Technology, Finland
Wei Fan	IBM T.J.Watson Research, USA
Paolo Frasconi	Università degli Studi di Firenze, Italy
João Gama	University of Porto, Portugal
Gemma C. Garriga	INRIA Lille Nord Europe, France
Claudio Gentile	Università dell'Insubria, Italy
Aristides Gionis	Yahoo! Research Barcelona, Spain
Geoff Holmes	University of Waikato, New Zealand
Eyke Hüllermeier	Philipps-Universität Marburg, Germany
George Karypis	University of Minnesota, USA
Kristian Kersting	University of Bonn, Germany
Joost Kok	University of Leiden, the Netherlands
James Kwok	Hong Kong University of Science and Technology, China
Bing Liu	University of Illinois, USA
Marie-Francine Moens	Katholieke Universiteit Leuven, Belgium
Alessandro Moschitti	University of Trento, Italy
Mahesan Niranjan	University of Southampton, UK
Dino Pedreschi	Università di Pisa, Italy
Jian Pei	Simon Fraser University, Canada
Bernhard Pfahringer	University of Waikato, New Zealand
Teemu Roos	University of Helsinki, Finland
Arno Siebes	University of Utrecht, the Netherlands
Myra Spiliopoulou	University of Magdeburg, Germany

Hannu Toivonen	University of Helsinki, Finland
Luis Torgo	University of Porto, Portugal
Jean-Philippe Vert	Mines ParisTech & Curie Institute, France
Stefan Wrobel	University of Bonn & Fraunhofer, Germany

Programme Committee

Naoki Abe	Stephane Canu	Brian Gallagher
Nitin Agarwal	Olivier Cappé	Patrick Gallinari
Fabio Aioli	Xavier Carreras	Eric Gaussier
Florence d'Alche-Buc	Andre Carvalho	Ricard Gavaldà
Aris Anagnostopoulos	Alexandra Carvalho	Peter Geibel
Gennady Andrienko	Michelangelo Ceci	Pierre Geurts
Ana Paula Appel	Tania Cerquitelli	Mohammad Ghavamzadeh
Marta Arias	Hong Cheng	Fosca Giannotti
Ira Assent	Weiwei Cheng	Rémi Gilleron
Martin Atzmueller	Jesús Cid-Sueiro	Christophe G.
Bart Baesens	Frans Coenen	Giraud-Carrier
José Balcázar	Fabrizio Costa	Aris Gkoulalas-Divanis
Elena Baralis	James Cussens	Bart Goethals
Shai Ben David	Alfredo Cuzzocrea	Marco Gori
Bettina Berendt	Jesse Davis	Henrik Grosskreutz
Michele Berlingerio	Krzysztof Dembczyński	Steve R. Gunn
Michael W. Berry	Janez Demšar	Stephan Günnemann
Michael R. Berthold	Christian Desrosiers	Dimitrios Gunopulos
Albert Bifet	Tom Diethe	Jiawei Han
Misha Bilenko	Kurt Driessens	Edwin Hancock
Mustafa Bilgic	Chris Drummond	Mohammad Hasan
Paolo Boldi	Pierre Dupont	Mark Herbster
Mario Boley	Saso Dzeroski	José Hernández-Orallo
Christian Borgelt	Floriana Esposito	Colin de la Higuera
Henrik Boström	A. Fazel Famili	Alexander Hinneburg
Stephane Boucheron	Nicola Fanizzi	Frank Hoepfner
Remco R. Bouckaert	Tom Fawcett	Jaakko Hollmén
Anne-Laure Boulesteix	Ad Feelders	Tamas Horvath
Jean-Francois Boulicaut	Xiaoli Z. Fern	Andreas Hotho
Marc Boullé	Cèsar Ferri	Minqing Hu
Pavel Brazdil	Daan Fierens	Ming Hua
Ulf Brefeld	Patrik Floréen	Ramon Huerta
Sebastien Bubeck	George Forman	Georgiana Ifrim
Wray Buntine	Blaž Fortuna	Nathalie Japkowicz
Tiberio Caetano	Elisa Fromont	Matti Järvisalo
Deng Cai	Johannes Fürnkranz	Alexandros Kalousis
Toon Calders	Mohamed Gaber	Hilbert Kappen
Colin Campbell	Thomas Gärtner	Hillol Kargupta

Panagiotis Karras	Olfa Nasraoui	Ulrich Rueckert
Hisashi Kashima	Sriraam Natarajan	Stefan Rüping
Samuel Kaski	Jennifer Neville	Maytal Saar-Tsechansky
Latifur Khan	Yizhao Ni	Lorenza Saitta
Marius Kloft	Siegfried Nijssen	Scott Sanner
Mikko Koivisto	Keith Noto	Vítor Santos Costa
Tamara G. Kolda	Andreas Nürnberger	Raul Santos-Rodriguez
Petri Kontkanen	Guillaume Obozinski	Tobias Scheffer
Walter A. Kosters	Francesco Orabona	Lars Schmidt-Thieme
Samory Kpotufe	Salvatore Orlando	Dale Schuurmans
Stefan Kramer	Gerhard Paaß	Michèle Sebag
Shonali Krishnaswamy	Tapio Pahikkala	Thomas Seidl
Nicolas Lachiche	George Paliouras	Ricardo Silva
Nada Lavrač	Spiros Papadimitriou	Andrzej Skowron
Matthijs Van Leeuwen	Panagiotis Papapetrou	Kevin Small
Jiuyong Li	Emilio	Carlos Soares
Tao Li	Parrado-Hernandez	Richard Socher
Xiaoli Li	Srinivasan Parthasarathy	Maarten van Someren
Jefrey Lijffijt	Andrea Passerini	Mauro Sozio
Aristidis Likas	Mykola Pechenizkiy	Alessandro Sperduti
Charles Ling	Marcello Pelillo	Masashi Sugiyama
Marco Lippi	Jaakko Peltonen	Jimeng Sun
Francesca A. Lisi	Pedro Pereira Rodrigues	Johan Suykens
Xiaohui Liu	Fernando Perez-Cruz	Einoshin Suzuki
Corrado Loglisci	Gregory	Sandor Szedmak
Daniel Lowd	Piatetsky-Shapiro	Prasad Tadepalli
Sofus A. Macskassy	Enric Plaza	Andrea Tagarelli
Donato Malerba	Massimiliano Pontil	Pang-Ning Tan
Giuseppe Manco	Ronaldo C. Prati	Lei Tang
Dragos D Margineantu	Kai Puolamäki	Nikolaj Tatti
Stan Matwin	Chedy Raïssi	Evimaria Terzi
Dimitrios Mavroeidis	Alain Rakotomamonjy	Ivan Titov
Michael May	Liva Ralaivola	Ljupčo Todorovski
Mike Mayo	Naren Ramakrishnan	Hanghang Tong
Thorsten Meinl	Jan Ramon	Volker Tresp
Prem Melville	Huzefa Rangwala	Konstantin Tretyakov
Ernestina Menasalvas	Balaraman Ravindran	Ivor W. Tsang
Aditya Menon	Patricia Riddle	Panayiotis Tsaparas
Rosa Meo	Fabrizio Riguzzi	Grigorios Tsoumakas
Pauli Miettinen	Fabio Roli	Koji Tsuda
Dunja Mladenić	Lorenzo Rosasco	Alan Tucker
Katharina J. Morik	Volker Roth	Antti Ukkonen
Klaus-Robert Muller	Juho Rousu	Joaquin Vanschoren
Emmanuel Müller	Céline Rouveirol	Michalis Vazirgiannis
Ricardo Nanculef	Cynthia Rudin	Shankar Vembu

Herna L. Viktor	Hui Xiong	Min-Ling Zhang
Fabio Vitale	Zhao Xu	Mark Zhang
Christel Vrain	Jieping Ye	Ying Zhao
Jilles Vreeken	Mi-Yen Yeh	Zheng Zhao
Willem Waegeman	Philip Yu	Zhi-Hua Zhou
Jianyong Wang	Gerson Zaverucha	Arthur Zimek
Hongning Wang	Filip Železný	Indre Zliobaite
Liwei Wang	Dell Zhang	Jean-Daniel Zucker
Wei Wang	Kai Zhang	Blaž Zupan
Gerhard Widmer	Lei Zhang	

Demo Track Programme Committee

Omar Alonso	Arvind Hulgeri	Mykola Pechenizkiy
Steve Beitzel	Ralf Klinkenberg	Peter van der Putten
Paul Bennett	Michael Mampaey	Daniela Stojanova
Michael Berthold	Michael May	Grigorios Tsoumakas
Albert Bifet	Gabor Melli	Karsten Winkler
Francesco Bonchi	Gerard de Melo	Michael Witbrock
Christian Borgelt	Rosa Meo	
Jaakko Hollmén	Themis Palpanas	

Nectar Track Programme Committee

Jason Baldridge	David R. Hardoon	Burr Settles
Xavier Carreras	Kristian Kersting	Ivan Titov
Pádraig Cunningham	Roni Khardon	Gyorgy Turan
Marc Deisenroth	Christina Leslie	Jean-Philippe Vert
Kurt Driessens	David Page	Zhao Xu
Mohammad Ghavamzadeh	Liva Ralaivola	
Aristides Gionis	Steffen Rendle	

Additional Reviewers

Artur Abdullin	Battista Biggio	Wing Kwan Chan
Ildefons Magrans de Abril	Sam Blasiak	Anveshi Charuvaka
Claudia d'Amato	Brigitte Boden	Silvia Chiusano
Haitham B. Ammar	Janez Brank	Michele Coscia
Fabrizio Angiulli	Forrest Briggs	Dominik Dahlem
Josh Attenberg	Giulia Bruno	Xuan-Hong Dang
Mohamad Azar	Samuel Rota Bulò	Hongbo Deng
Luca Baldassarre	Luca Cagliero	Nicola Di Mauro
Martin Becker	Rui Camacho	Luca Didaci
Jaafar Ben-Abdallah	Hong Cao	Huyen T. Do
	Loic Cerf	Gauthier Doquire

Nikolaos Engonopoulos	Guy Lever	Giulio Rossetti
Chaosheng Fan	Lin Liu	Natali Ruchansky
Kai Fan	Grigorios Loukides	Patricia Iglesias Sánchez
Wei Feng	Cécile Low-Kam	Tanwishta Saha
Carlos Ferreira	Thomas Low	Esin Saka
Raphael Fonteneau	Frangiskos Malliaros	Antonio Bella Sanjuán
Benoît Frénay	Fernando	Jan Schlüter
Sergej Fries	Martinez-Plumed	Chun-Wei Seah
David Fuhry	Ida Mele	Wei Shen
Fabio Fumarola	João Mendes Moreira	Marcin Skowron
Victor Gabillon	Glauber Marcius	Eleftherios
Esther Galbrun	Menezes	Spyromitros-Xioufis
Shenghua Gao	Barbora Micenková	Tadej Stajner
Aurélien Garivier	Pasquale Minervini	Daniela Stojanova
Konstantinos Georgatzis	Joseph Modayil	Hongyu Su
Christos Giatsidis	Anna Monreale	Yizhou Sun
Tatiana Gossen	Tetsuro Morimura	Akiko Takeda
Quanquan Gu	James Neufeld	Frank W. Takes
Francesco Gullo	Minh Nhut Nguyen	Jafar Tanha
Manish Gupta	Maximilian Nickel	Claudio Taranto
Basheer Hawwash	Inna Novalija	Sep Thijssen
Jingrui He	Christopher Oßner	Stamatina Thomaidou
Todd Hester	Aline Marins Paes	Daniel Trabold
Xin Huang	Joni Pajarinen	Mitja Trampus
Yi Huang	Luca Pappalardo	Roberto Trasarti
Dino Ienco	Eric Paquet	Erik Tromp
Roberto Interdonato	Daniel Paurat	Yuta Tsuboi
Baptiste Jeudy	Yuanli Pei	Duygu Ucar
Lili Jiang	Ruggero G. Pensa	Michal Valko
Xueyan Jiang	Claudia Perlich	Ugo Vespier
Michael Kamp	Sergios Petridis	Silvia Villa
Emilie Kaufmann	Anja Pilz	Yana Volkovich
Fabian Keller	Gianvito Pio	Byron C. Wallace
Ryan Kiros	Cristiano Grijó Pitangui	Jun Wang
Julia Kiseleva	Marthinus Christoffel	Xiaodan Wang
Hannes Korte	du Plessis	Pascal Welke
Aris Kosmopoulos	Vid Podpečan	Makoto Yamada
Petra Kralj Novak	Chiara Pulice	Xintian Yang
Philipp Kranen	Miao Qiao	Jihang Ye
Hardy Kremer	M. Jose	Eric Yeh
Anastasia Krithara	Ramirez-Quintana	Guoxian Yu
Denis Krompass	Zeehasham Rasheed	Yaoliang Yu
Aggeliki Lazaridou	Irene Rodriguez Lujan	Elias Zavitsanos
Florian Lemmerich	Bernardino	Wei Zhang
Patrick Lessman	Romera-Paredes	Tingting Zhao

Table of Contents – Part I

Invited Talks

Machine Learning for Robotics	1
<i>Pieter Abbeel</i>	
Declarative Modeling for Machine Learning and Data Mining	2
<i>Luc De Raedt</i>	
Machine Learning Methods for Music Discovery and Recommendation	4
<i>Douglas Eck</i>	
Solving Problems with Visual Analytics: Challenges and Applications . . .	5
<i>Daniel Keim</i>	
Analyzing Text and Social Network Data with Probabilistic Models . . .	7
<i>Padhraic Smyth</i>	

Association Rules and Frequent Patterns

Discovering Descriptive Tile Trees: By Mining Optimal Geometric Subtiles	9
<i>Nikolaj Tatti and Jilles Vreeken</i>	
Efficient Discovery of Association Rules and Frequent Itemsets through Sampling with Tight Performance Guarantees	25
<i>Matteo Riondato and Eli Upfal</i>	
Smoothing Categorical Data	42
<i>Arno Siebes and René Kersten</i>	

Bayesian Learning and Graphical Models

An Experimental Comparison of Hybrid Algorithms for Bayesian Network Structure Learning	58
<i>Maxime Gasse, Alex Aussem, and Haytham Elghazel</i>	
Bayesian Network Classifiers with Reduced Precision Parameters	74
<i>Sebastian Tschiatschek, Peter Reinprecht, Manfred Mücke, and Franz Pernkopf</i>	
Combining Subjective Probabilities and Data in Training Markov Logic Networks	90
<i>Tivadar Pápai, Shalini Ghosh, and Henry Kautz</i>	

Score-Based Bayesian Skill Learning 106
Shengbo Guo, Scott Sanner, Thore Graepel, and Wray Buntine

Classification

A Note on Extending Generalization Bounds for Binary Large-Margin
Classifiers to Multiple Classes 122
Ürün Dogan, Tobias Glasmachers, and Christian Igel

Extension of the Rocchio Classification Method to Multi-modal
Categorization of Documents in Social Media 130
Amin Mantrach and Jean-Michel Renders

Label-Noise Robust Logistic Regression and Its Applications 143
Jakramate Bootkrajang and Ata Kabán

Sentiment Classification with Supervised Sequence Embedding 159
Dmitriy Bespalov, Yanjun Qi, Bing Bai, and Ali Shokoufandeh

The Bitvector Machine: A Fast and Robust Machine Learning
Algorithm for Non-linear Problems 175
Stefan Edelkamp and Martin Stommel

**Dimensionality Reduction, Feature Selection and
Extraction**

Embedding Monte Carlo Search of Features in Tree-Based Ensemble
Methods 191
Francis Maes, Pierre Geurts, and Louis Wehenkel

Hypergraph Spectra for Semi-supervised Feature Selection 207
Zhihong Zhang, Edwin R. Hancock, and Xiao Bai

Learning Neighborhoods for Metric Learning 223
Jun Wang, Adam Woznica, and Alexandros Kalousis

Massively Parallel Feature Selection: An Approach Based on Variance
Preservation 237
Zheng Zhao, James Cox, David Duling, and Warren Sarle

PCA, Eigenvector Localization and Clustering for Side-Channel Attacks
on Cryptographic Hardware Devices 253
*Dimitrios Mavroeidis, Lejla Batina, Twan van Laarhoven, and
Elena Marchiori*

Distance-Based Methods and Kernels

Classifying Stem Cell Differentiation Images by Information Distance . . .	269
<i>Xianglilan Zhang, Hongnan Wang, Tony J. Collins, Zhiqiang Luo, and Ming Li</i>	
Distance Metric Learning Revisited	283
<i>Qiong Cao, Yiming Ying, and Peng Li</i>	
Geodesic Analysis on the Gaussian RKHS Hypersphere	299
<i>Nicolas Courty, Thomas Burger, and Pierre-François Marteau</i>	

Ensemble Methods

Boosting Nearest Neighbors for the Efficient Estimation of Posteriors . . .	314
<i>Roberto D’Ambrosio, Richard Nock, Wafa Bel Haj Ali, Frank Nielsen, and Michel Barlaud</i>	
Diversity Regularized Ensemble Pruning	330
<i>Nan Li, Yang Yu, and Zhi-Hua Zhou</i>	
Ensembles on Random Patches	346
<i>Gilles Louppe and Pierre Geurts</i>	

Graph and Tree Mining

An Efficiently Computable Support Measure for Frequent Subgraph Pattern Mining	362
<i>Yuyi Wang and Jan Ramon</i>	
Efficient Graph Kernels by Randomization	378
<i>Marion Neumann, Novi Patricia, Roman Garnett, and Kristian Kersting</i>	
Graph Mining for Object Tracking in Videos	394
<i>Fabien Diot, Elisa Fromont, Baptiste Jeudy, Emmanuel Marilly, and Olivier Martinot</i>	
Hypergraph Learning with Hyperedge Expansion	410
<i>Li Pu and Boi Faltings</i>	
Nearly Exact Mining of Frequent Trees in Large Networks	426
<i>Ashraf M. Kibriya and Jan Ramon</i>	
Reachability Analysis and Modeling of Dynamic Event Networks	442
<i>Kathy Macropol and Ambuj Singh</i>	

Large-Scale, Distributed and Parallel Mining and Learning

CC-MR – Finding Connected Components in Huge Graphs with MapReduce	458
<i>Thomas Seidl, Brigitte Boden, and Sergej Fries</i>	
Fast Near Neighbor Search in High-Dimensional Binary Data	474
<i>Anshumali Shrivastava and Ping Li</i>	
Fully Sparse Topic Models	490
<i>Khoat Than and Tu Bao Ho</i>	
Learning Compact Class Codes for Fast Inference in Large Multi Class Classification	506
<i>M. Cissé, T. Artières, and Patrick Gallinari</i>	
ParCube: Sparse Parallelizable Tensor Decompositions	521
<i>Evangelos E. Papalexakis, Christos Faloutsos, and Nicholas D. Sidiropoulos</i>	
Stochastic Coordinate Descent Methods for Regularized Smooth and Nonsmooth Losses	537
<i>Qing Tao, Kang Kong, Dejun Chu, and Gaowei Wu</i>	
Sublinear Algorithms for Penalized Logistic Regression in Massive Datasets	553
<i>Haoruo Peng, Zhengyu Wang, Edward Y. Chang, Shuchang Zhou, and Zhihua Zhang</i>	

Multi-Relational Mining and Learning

Author Name Disambiguation Using a New Categorical Distribution Similarity	569
<i>Shaohua Li, Gao Cong, and Chunyan Miao</i>	
Lifted Online Training of Relational Models with Stochastic Gradient Methods	585
<i>Babak Ahmadi, Kristian Kersting, and Sriraam Natarajan</i>	
Scalable Relation Prediction Exploiting Both Intrarelatonal Correlation and Contextual Information	601
<i>Xueyan Jiang, Volker Tresp, Yi Huang, Maximilian Nickel, and Hans-Peter Kriegel</i>	
Relational Differential Prediction	617
<i>Houssam Nassif, Vítor Santos Costa, Elizabeth S. Burnside, and David Page</i>	

Multi-Task Learning

Efficient Training of Graph-Regularized Multitask SVMs	633
<i>Christian Widmer, Marius Kloft, Nico Görnitz, and Gunnar Rätsch</i>	
Geometry Preserving Multi-task Metric Learning	648
<i>Peipei Yang, Kaizhu Huang, and Cheng-Lin Liu</i>	
Learning and Inference in Probabilistic Classifier Chains with Beam Search	665
<i>Abhishek Kumar, Shankar Vembu, Aditya Krishna Menon, and Charles Elkan</i>	
Learning Multiple Tasks with Boosted Decision Trees	681
<i>Jean Baptiste Faddoul, Boris Chidlovskii, Rémi Gilleron, and Fabien Torre</i>	
Multi-Task Boosting by Exploiting Task Relationships	697
<i>Yu Zhang and Dit-Yan Yeung</i>	
Sparse Gaussian Processes for Multi-task Learning	711
<i>Yuyang Wang and Roni Khardon</i>	

Natural Language Processing

Collective Information Extraction with Context-Specific Consistencies	728
<i>Peter Kluegl, Martin Toepfer, Florian Lemmerich, Andreas Hotho, and Frank Puppe</i>	
Supervised Learning of Semantic Relatedness	744
<i>Ran El-Yaniv and David Yanay</i>	
Unsupervised Bayesian Part of Speech Inference with Particle Gibbs	760
<i>Gregory Dubbin and Phil Blunsom</i>	
WikiSent: Weakly Supervised Sentiment Analysis through Extractive Summarization with Wikipedia	774
<i>Subhabrata Mukherjee and Pushpak Bhattacharyya</i>	

Online Learning and Data Streams

Adaptive Two-View Online Learning for Math Topic Classification	794
<i>Tam T. Nguyen, Kuiyu Chang, and Siu Cheung Hui</i>	
BDUOL: Double Updating Online Learning on a Fixed Budget	810
<i>Peilin Zhao and Steven C.H. Hoi</i>	

Handling Time Changing Data with Adaptive Very Fast Decision Rules	827
<i>Petr Kosina and João Gama</i>	
Improved Counter Based Algorithms for Frequent Pairs Mining in Transactional Data Streams	843
<i>Konstantin Kutzkov</i>	
Mirror Descent for Metric Learning: A Unified Approach	859
<i>Gautam Kunapuli and Jude Shavlik</i>	
Author Index	875

Table of Contents – Part II

Privacy and Security

<i>AUDIO</i> : An Integrity <i>A</i> uditing Framework of Outlier-Mining-as-a-Service Systems	1
<i>Ruilin Liu, Hui (Wendy) Wang, Anna Monreale, Dino Pedreschi, Fosca Giannotti, and Wenge Guo</i>	
Differentially Private Projected Histograms: Construction and Use for Prediction	19
<i>Staal A. Vinterbo</i>	
Fairness-Aware Classifier with Prejudice Remover Regularizer	35
<i>Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma</i>	

Rankings and Recommendations

A Live Comparison of Methods for Personalized Article Recommendation at Forbes.com	51
<i>Evan Kirshenbaum, George Forman, and Michael Dugan</i>	
Fast ALS-Based Tensor Factorization for Context-Aware Recommendation from Implicit Feedback	67
<i>Balázs Hidasi and Domonkos Tikk</i>	
Probability Estimation for Multi-class Classification Based on Label Ranking	83
<i>Weiwei Cheng and Eyke Hüllermeier</i>	

Reinforcement Learning and Planning

Adaptive Planning for Markov Decision Processes with Uncertain Transition Models via Incremental Feature Dependency Discovery	99
<i>N. Kemal Ure, Alborz Geramifard, Girish Chowdhary, and Jonathan P. How</i>	
APRIL: Active Preference Learning-Based Reinforcement Learning	116
<i>Riad Akrouf, Marc Schoenauer, and Michèle Sebag</i>	
Autonomous Data-Driven Decision-Making in Smart Electricity Markets	132
<i>Markus Peters, Wolfgang Ketter, Maytal Saar-Tsechansky, and John Collins</i>	

Bayesian Nonparametric Inverse Reinforcement Learning	148
<i>Bernard Michini and Jonathan P. How</i>	
Bootstrapping Monte Carlo Tree Search with an Imperfect Heuristic	164
<i>Truong-Huy Dinh Nguyen, Wee-Sun Lee, and Tze-Yun Leong</i>	
Fast Reinforcement Learning with Large Action Sets Using Error-Correcting Output Codes for MDP Factorization	180
<i>Gabriel Dulac-Arnold, Ludovic Denoyer, Philippe Preux, and Patrick Gallinari</i>	
Learning Policies for Battery Usage Optimization in Electric Vehicles	195
<i>Stefano Ermon, Yexiang Xue, Carla Gomes, and Bart Selman</i>	
Policy Iteration Based on a Larned Transition Model	211
<i>Vivek Ramavajjala and Charles Elkan</i>	
Structured Apprenticeship Learning	227
<i>Abdeslam Boularias, Oliver Krömer, and Jan Peters</i>	

Rule Mining and Subgroup Discovery

A Bayesian Approach for Classification Rule Mining in Quantitative Databases	243
<i>Dominique Gay and Marc Boullé</i>	
A Bayesian Scoring Technique for Mining Predictive and Non-Spurious Rules	260
<i>Iyad Batal, Gregory Cooper, and Milos Hauskrecht</i>	
Generic Pattern Trees for Exhaustive Exceptional Model Mining	277
<i>Florian Lemmerich, Martin Becker, and Martin Atzmueller</i>	

Semi-Supervised and Transductive Learning

Bidirectional Semi-supervised Learning with Graphs	293
<i>Tomoharu Iwata and Kevin Duh</i>	
Coupled Bayesian Sets Algorithm for Semi-supervised Learning and Information Extraction	307
<i>Saurabh Verma and Estevam R. Hruschka Jr.</i>	
Graph-Based Transduction with Confidence	323
<i>Matan Orbach and Koby Crammer</i>	
Maximum Consistency Preferential Random Walks	339
<i>Deguang Kong and Chris Ding</i>	

Semi-supervised Multi-label Classification: A Simultaneous Large-Margin, Subspace Learning Approach	355
<i>Yuhong Guo and Dale Schuurmans</i>	

Sensor Data

MDL-Based Analysis of Time Series at Multiple Time-Scales	371
<i>Ugo Vespier, Arno Knobbe, Siegfried Nijssen, and Joaquin Vanschoren</i>	
Separable Approximate Optimization of Support Vector Machines for Distributed Sensing	387
<i>Sangkyun Lee, Marco Stolpe, and Katharina Morik</i>	
Unsupervised Inference of Auditory Attention from Biosensors	403
<i>Melih Kandemir, Arto Klami, Akos Vetek, and Samuel Kaski</i>	

Sequence and String Mining

A Family of Feed-Forward Models for Protein Sequence Classification	419
<i>Sam Blasiak, Huzefa Rangwala, and Kathryn B. Laskey</i>	
General Algorithms for Mining Closed Flexible Patterns under Various Equivalence Relations	435
<i>Tomohiro I, Yuki Enokuma, Hideo Bannai, and Masayuki Takeda</i>	
Size Matters: Finding the Most Informative Set of Window Lengths	451
<i>Jefrey Lijffijt, Panagiotis Papapetrou, and Kai Puolamäki</i>	

Social Network Mining

Discovering Links among Social Networks	467
<i>Francesco Buccafurri, Gianluca Lax, Antonino Nocera, and Domenico Ursino</i>	
Efficient Bi-objective Team Formation in Social Networks	483
<i>Mehdi Kargar, Aijun An, and Morteza Zihayat</i>	
Feature-Enhanced Probabilistic Models for Diffusion Network Inference	499
<i>Liaoruo Wang, Stefano Ermon, and John E. Hopcroft</i>	
Influence Spread in Large-Scale Social Networks – A Belief Propagation Approach	515
<i>Huy Nguyen and Rong Zheng</i>	

Location Affiliation Networks: Bonding Social and Spatial Information	531
<i>Konstantinos Pelechrinis and Prashant Krishnamurthy</i>	
On Approximation of Real-World Influence Spread	548
<i>Yu Yang, Enhong Chen, Qi Liu, Biao Xiang, Tong Xu, and Shafqat Ali Shad</i>	
Opinion Formation by Voter Model with Temporal Decay Dynamics	565
<i>Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda</i>	
Viral Marketing for Product Cross-Sell through Social Networks	581
<i>Ramasuri Narayanam and Amit A. Nanavati</i>	
Which Topic Will You Follow?	597
<i>Deqing Yang, Yanghua Xiao, Bo Xu, Hanghang Tong, Wei Wang, and Sheng Huang</i>	

Spatial and Geographical Data Mining

Inferring Geographic Coincidence in Ephemeral Social Networks	613
<i>Honglei Zhuang, Alvin Chin, Sen Wu, Wei Wang, Xia Wang, and Jie Tang</i>	
Pedestrian Quantity Estimation with Trajectory Patterns	629
<i>Thomas Liebig, Zhao Xu, Michael May, and Stefan Wrobel</i>	
Socioscope: Spatio-temporal Signal Recovery from Social Media	644
<i>Jun-Ming Xu, Aniruddha Bhargava, Robert Nowak, and Xiaojin Zhu</i>	

Statistical Methods and Evaluation

A Framework for Evaluating the Smoothness of Data-Mining Results . . .	660
<i>Gaurav Misra, Behzad Golshan, and Evimaria Terzi</i>	
Active Evaluation of Ranking Functions Based on Graded Relevance . . .	676
<i>Christoph Sawade, Steffen Bickel, Timo von Oertzen, Tobias Scheffer, and Niels Landwehr</i>	

Time Series and Temporal Data Mining

Community Trend Outlier Detection Using Soft Temporal Pattern Mining	692
<i>Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han</i>	
Data Structures for Detecting Rare Variations in Time Series	709
<i>Caio Valentim, Eduardo S. Laber, and David Sotelo</i>	

Invariant Time-Series Classification	725
<i>Josif Grabocka, Alexandros Nanopoulos, and Lars Schmidt-Thieme</i>	

Learning Bi-clustered Vector Autoregressive Models	741
<i>Tzu-Kuo Huang and Jeff Schneider</i>	

Transfer Learning

Discriminative Factor Alignment across Heterogeneous Feature Space . . .	757
<i>Fangwei Hu, Tianqi Chen, Nathan N. Liu, Qiang Yang, and Yong Yu</i>	

Learning to Perceive Two-Dimensional Displays Using Probabilistic Grammars	773
<i>Nan Li, William W. Cohen, and Kenneth R. Koedinger</i>	

Transfer Spectral Clustering	789
<i>Wenhao Jiang and Fu-lai Chung</i>	

System Demonstrations Track

An Aspect-Lexicon Creation and Evaluation Tool for Sentiment Analysis Researchers	804
<i>Mus'ab Husaini, Ahmet Koçyiğit, Dilek Tapucu, Berrin Yanikoglu, and Yücel Saygin</i>	

Association Rule Mining Following the Web Search Paradigm	808
<i>Radek Škrabal, Milan Šimůnek, Stanislav Vojříř, Andrej Hazucha, Tomáš Marek, David Chudán, and Tomáš Kliegr</i>	

ASV Monitor: Creating Comparability of Machine Learning Methods for Content Analysis	812
<i>Andreas Nickler, Patrick Jähnichen, and Gerhard Heyer</i>	

CloudFlows: A Cloud Based Scientific Workflow Platform	816
<i>Janez Kranjc, Vid Podpečan, and Nada Lavrač</i>	

Extracting Trajectories through an Efficient and Unifying Spatio-temporal Pattern Mining System	820
<i>Phan Nhat Hai, Dino Ienco, Pascal Poncelet, and Maguelonne Teisseire</i>	

Knowledge Discovery through Symbolic Regression with HeuristicLab	824
<i>Gabriel Kronberger, Stefan Wagner, Michael Kommenda, Andreas Beham, Andreas Scheibenpflug, and Michael Affenzeller</i>	

<i>OutRules</i> : A Framework for Outlier Descriptions in Multiple Context Spaces	828
<i>Emmanuel Müller, Fabian Keller, Sebastian Blanc, and Klemens Böhm</i>	
Scientific Workflow Management with ADAMS	833
<i>Peter Reutemann and Joaquin Vanschoren</i>	
TopicExplorer: Exploring Document Collections with Topic Models	838
<i>Alexander Hinneburg, Rico Preiss, and René Schröder</i>	
VIKAMINE – Open-Source: Subgroup Discovery, Pattern Mining, and Analytics	842
<i>Martin Atzmueller and Florian Lemmerich</i>	
Nectar Track	
Learning Submodular Functions	846
<i>Maria-Florina Balcan and Nicholas J.A. Harvey</i>	
Matrix Factorization as Search	850
<i>Kristian Kersting, Christian Bauckhage, Christian Thureau, and Mirwaes Wahabzada</i>	
Metal Binding in Proteins: Machine Learning Complements X-Ray Absorption Spectroscopy	854
<i>Marco Lippi, Andrea Passerini, Marco Punta, and Paolo Frasconi</i>	
Modelling Input Varying Correlations between Multiple Responses	858
<i>Andrew Gordon Wilson and Zoubin Ghahramani</i>	
Author Index	863