

Toon Calders Floriana Esposito
Eyke Hüllermeier Rosa Meo (Eds.)

LNAI 8724

Machine Learning and Knowledge Discovery in Databases

European Conference, ECML PKDD 2014
Nancy, France, September 15–19, 2014
Proceedings, Part I

1
Part I

 Springer

Lecture Notes in Artificial Intelligence 8724

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

Toon Calders Floriana Esposito
Eyke Hüllermeier Rosa Meo (Eds.)

Machine Learning and Knowledge Discovery in Databases

European Conference, ECML PKDD 2014
Nancy, France, September 15-19, 2014
Proceedings, Part I



Springer

Volume Editors

Toon Calders

Université Libre de Bruxelles, Faculty of Applied Sciences

Department of Computer and Decision Engineering

Av. F. Roosevelt, CP 165/15, 1050 Brussels, Belgium

E-mail: toon.calders@ulb.ac.be

Floriana Esposito

Università degli Studi "Aldo Moro", Dipartimento di Informatica

via Orabona 4, 70125 Bari, Italy

E-mail: floriana.esposito@uniba.it

Eyke Hüllermeier

Universität Paderborn, Department of Computer Science

Warburger Str. 100, 33098 Paderborn, Germany

E-mail: eyke@upb.de

Rosa Meo

Università degli Studi di Torino, Dipartimento di Informatica

Corso Svizzera 185, 10149 Torino, Italy

E-mail: meo@di.unito.it

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-662-44847-2

e-ISBN 978-3-662-44848-9

DOI 10.1007/978-3-662-44848-9

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014948041

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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

The European Conferences on Machine Learning (ECML) and on Principles and Practice of Knowledge Discovery in Data Bases (PKDD) have been organized jointly since 2001, after some years of mutual independence. Going one step further, the two conferences were merged into a single one in 2008, and these are the proceedings of the 2014 edition of ECML/PKDD. Today, this conference is a world-wide leading scientific event. It aims at further exploiting the synergies between the two scientific fields, focusing on the development and employment of methods and tools capable of solving real-life problems.

ECML PKDD 2014 was held in Nancy, France, during September 15–19, co-located with ILP 2014, the premier international forum on logic-based and relational learning. The two conferences were organized by Inria Nancy Grand Est with support from LORIA, a joint research unit of CNRS, Inria, and Université de Lorraine.

Continuing the tradition, ECML/PKDD 2014 combined an extensive technical program with a demo track and an industrial track. Recently, the so-called Nectar track was added, focusing on the latest high-quality interdisciplinary research results in all areas related to machine learning and knowledge discovery in databases. Moreover, the conference program included a discovery challenge, a variety of workshops, and many tutorials.

The main technical program included five plenary talks by invited speakers, namely, Charu Aggarwal, Francis Bach, Lise Getoor, Tie-Yan Liu, and Raymond Ng, while four invited speakers contributed to the industrial track: George Hébrail (EDF Lab), Alexandre Cotarmanac’h (Twenga), Arthur Von Eschen (Activision Publishing Inc.) and Mike Bodkin (Evotec Ltd.).

The discovery challenge focused on “Neural Connectomics and on Predictive Web Analytics” this year. Fifteen workshops were held, providing an opportunity to discuss current topics in a small and interactive atmosphere: Dynamic Networks and Knowledge Discovery, Interactions Between Data Mining and Natural Language Processing, Mining Ubiquitous and Social Environments, Statistically Sound Data Mining, Machine Learning for Urban Sensor Data, Multi-Target Prediction, Representation Learning, Neural Connectomics: From Imaging to Connectivity, Data Analytics for Renewable Energy Integration, Linked Data for Knowledge Discovery, New Frontiers in Mining Complex Patterns, Experimental Economics and Machine Learning, Learning with Multiple Views: Applications to Computer Vision and Multimedia Mining, Generalization and Reuse of Machine Learning Models over Multiple Contexts, and Predictive Web Analytics.

Nine tutorials were included in the conference program, providing a comprehensive introduction to core techniques and areas of interest for the scientific community: Medical Mining for Clinical Knowledge Discovery, Patterns in Noisy and Multidimensional Relations and Graphs, The Pervasiveness of

Machine Learning in Omics Science, Conformal Predictions for Reliable Machine Learning, The Lunch Is Never Free: How Information Theory, MDL, and Statistics are Connected, Information Theoretic Methods in Data Mining, Machine Learning with Analogical Proportions, Preference Learning Problems, and Deep Learning.

The main track received 481 paper submissions, of which 115 were accepted. Such a high volume of scientific work required a tremendous effort by the area chairs, Program Committee members, and many additional reviewers. We managed to collect three highly qualified independent reviews per paper and one additional overall input from one of the area chairs. Papers were evaluated on the basis of their relevance to the conference, their scientific contribution, rigor and correctness, the quality of presentation and reproducibility of experiments. As a separate organization, the demo track received 24 and the Nectar track 23 paper submissions.

For the second time, the conference used a double submission model: next to the regular conference track, papers submitted to the Springer journals *Machine Learning* (MACH) and *Data Mining and Knowledge Discovery* (DAMI) were considered for presentation in the conference. These papers were submitted to the ECML/PKDD 2014 special issue of the respective journals, and underwent the normal editorial process of these journals. Those papers accepted for the of these journals were assigned a presentation slot at the ECML/PKDD 2014 conference. A total of 107 original manuscripts were submitted to the journal track, 15 were accepted in DAMI or MACH and were scheduled for presentation at the conference. Overall, this resulted in a number of 588 submissions, of which 130 were selected for presentation at the conference, making an overall acceptance rate of about 22%.

These proceedings of the ECML/PKDD 2014 conference contain the full papers of the contributions presented in the main technical track, abstracts of the invited talks and short papers describing the demonstrations, and the Nectar papers. First of all, we would like to express our gratitude to the general chairs of the conference, Amedeo Napoli and Chedy Raïssi, as well as to all members of the Organizing Committee, for managing this event in a very competent and professional way. In particular, we thank the demo, workshop, industrial, and Nectar track chairs. Special thanks go to the proceedings chairs, Élisabeth Fromont, Stefano Ferilli and Pascal Poncelet, for the hard work of putting these proceedings together. We thank the tutorial chairs, the Discovery Challenge organizers and all the people involved in the conference, who worked hard for its success. Last but not least, we would like to sincerely thank the authors for submitting their work to the conference and the reviewers and area chairs for their tremendous effort in guaranteeing the quality of the reviewing process, thereby improving the quality of these proceedings.

July 2014

Toon Calders
Floriana Esposito
Eyke Hüllermeier
Rosa Meo

Organization

ECML/PKDD 2014 Organization

Conference Co-chairs

| | |
|---------------|-------------------------------------|
| Amedeo Napoli | Inria Nancy Grand Est/LORIA, France |
| Chedy Raïssi | Inria Nancy Grand Est/LORIA, France |

Program Co-chairs

| | |
|-------------------|--|
| Toon Calders | Université Libre de Bruxelles, Belgium |
| Floriana Esposito | University of Bari, Italy |
| Eyke Hüllermeier | University of Paderborn, Germany |
| Rosa Meo | University of Turin, Italy |

Local Organization Co-chairs

| | |
|-----------------------|-------------------------------|
| Anne-Lise Charbonnier | Inria Nancy Grand Est, France |
| Louisa Touioui | Inria Nancy Grand Est, France |

Awards Committee Chairs

| | |
|--------------------|--|
| Johannes Fürnkranz | Technical University of Darmstadt, Germany |
| Katharina Morik | University of Dortmund, Germany |

Workshop Chairs

| | |
|-------------------|--------------------|
| Bettina Berendt | KU Leuven, Belgium |
| Patrick Gallinari | LIP6 Paris, France |

Tutorial Chairs

| | |
|------------------|----------------------------------|
| Céline Rouveirol | University of Paris-Nord, France |
| Céline Robardet | University of Lyon, France |

Demonstration Chairs

| | |
|-------------------|----------------------------------|
| Ricard Gavaldà | UPC Barcelona, Spain |
| Myra Spiliopoulou | University of Magdeburg, Germany |

Publicity Chairs

Stefano Ferilli
Pauli Miettinen

University of Bari, Italy
Max-Planck-Institut, Germany

Panel Chairs

Jose Balcazar
Sergei O. Kuznetsov

UPC Barcelona, Spain
HSE Moscow, Russia

Industrial Chairs

Michael Berthold
Marc Boullé

University of Konstanz, Germany
Orange Labs, France

PhD Chairs

Bruno Crémilleux
Radim Belohlavek

University of Caen, France
University of Olomouc, Czech Republic

Nectar Track Chairs

Evimaria Terzi
Pierre Geurts

Boston University, USA
University of Liège, Belgium

Sponsorship Chairs

Francesco Bonchi
Jilles Vreeken

Yahoo ! Research Barcelona, Spain
Saarland University/Max-Planck-Institut,
Germany

Proceedings Chairs

Pascal Poncelet
Élisa Fromont
Stefano Ferilli

University of Montpellier, France
University of Saint Etienne, France
University of Bari, Italy

EMCL PKDD Steering Committee

Fosca Giannotti
Michèle Sebag
Francesco Bonchi
Hendrik Blockeel

University of Pisa, Italy
Université Paris Sud, France
Yahoo! Research Barcelona, Spain
KU Leuven, Belgium and Leiden University,
The Netherlands

| | |
|-----------------------|---|
| Katharina Morik | University of Dortmund, Germany |
| Tobias Scheffer | University of Potsdam, Germany |
| Arno Siebes | Utrecht University, The Netherlands |
| Dimitrios Gunopulos | University of Athens, Greece |
| Michalis Vazirgiannis | École Polytechnique, France |
| Donato Malerba | University of Bari, Italy |
| Peter Flach | University of Bristol, UK |
| Tijl De Bie | University of Bristol, UK |
| Nello Cristianini | University of Bristol, UK |
| Filip Železný | Czech Technical University in Prague, Czech Republic |
| Siegfried Nijssen | LIACS, Leiden University, The Netherlands |
| Kristian Kersting | Technical University of Dortmund, Germany |

Area Chairs

| | |
|----------------------|---|
| Hendrik Blockeel | KU Leuven, Belgium |
| Henrik Boström | Stockholm University, Sweden |
| Ian Davidson | University of California, Davis, USA |
| Luc De Raedt | KU Leuven, Belgium |
| Janez Demšar | University of Ljubljana, Slovenia |
| Alan Fern | Oregon State University, USA |
| Peter Flach | University of Bristol, UK |
| Johannes Fürnkranz | TU Darmstadt, Germany |
| Thomas Gärtner | University of Bonn and Fraunhofer IAIS, Germany |
| João Gama | University of Porto, Portugal |
| Aristides Gionis | Aalto University, Finland |
| Bart Goethals | University of Antwerp, Belgium |
| Andreas Hotho | University of Würzburg, Germany |
| Manfred Jaeger | Aalborg University, Denmark |
| Thorsten Joachims | Cornell University, USA |
| Kristian Kersting | Technical University of Dortmund, Germany |
| Stefan Kramer | University of Mainz, Germany |
| Donato Malerba | University of Bari, Italy |
| Stan Matwin | Dalhousie University, Canada |
| Pauli Miettinen | Max-Planck-Institut, Germany |
| Dunja Mladenić | Jozef Stefan Institute, Slovenia |
| Marie-Francine Moens | KU Leuven, Belgium |
| Bernhard Pfahringer | University of Waikato, New Zealand |
| Thomas Seidl | RWTH Aachen University, Germany |
| Arno Siebes | Utrecht University, The Netherlands |
| Myra Spiliopoulou | Magdeburg University, Germany |
| Jean-Philippe Vert | Mines ParisTech, France |
| Jilles Vreeken | Max-Planck-Institut and Saarland University, Germany |

Marco Wiering
Stefan Wrobel

University of Groningen, The Netherlands
University of Bonn & Fraunhofer IAIS,
Germany

Program Committee

| | | |
|-------------------------|------------------------|------------------------|
| Foto Afrati | Wray Buntine | Wei Ding |
| Leman Akoglu | Robert Busa-Fekete | Ying Ding |
| Mehmet Sabih Aksoy | Toon Calders | Stephan Doerfel |
| Mohammad Al Hasan | Rui Camacho | Janardhan Rao Doppa |
| Omar Alonso | Longbing Cao | Chris Drummond |
| Aijun An | Andre Carvalho | Devdatt Dubhashi |
| Aris Anagnostopoulos | Francisco Casacuberta | Ines Dutra |
| Annalisa Appice | Michelangelo Ceci | Sašo Džeroski |
| Marta Arias | Loic Cerf | Tapio Elomaa |
| Hiroki Arimura | Tania Cerquitelli | Roberto Esposito |
| Ira Assent | Sharma Chakravarthy | Ines Faerber |
| Martin Atzmüller | Keith Chan | Hadi Fanaee-Tork |
| Chloe-Agathe Azencott | Duen Horng Chau | Nicola Fanizzi |
| Antonio Bahamonde | Sanjay Chawla | Elaine Faria |
| James Bailey | Keke Chen | Fabio Fassetti |
| Elena Baralis | Ling Chen | Hakan |
| Daniel Barbara' | Weiwei Cheng | Ferhatosmanoglou |
| Christian Bauckhage | Silvia Chiusano | Stefano Ferilli |
| Roberto Bayardo | Vassilis Christophides | Carlos Ferreira |
| Aurelien Bellet | Frans Coenen | Cèsar Ferri |
| Radim Belohlavek | Fabrizio Costa | Jose Fonollosa |
| Andras Benczur | Bruno Cremilleux | Eibe Frank |
| Klaus Berberich | Tom Croonenborghs | Antonino Freno |
| Bettina Berendt | Boris Cule | Élisa Fromont |
| Michele Berlingerio | Tomaz Curk | Fabio Fumarola |
| Indrajit Bhattacharya | James Cussens | Patrick Gallinari |
| Marenglen Biba | Maria Damiani | Jing Gao |
| Albert Bifet | Jesse Davis | Byron Gao |
| Enrico Blanzieri | Martine De Cock | Roman Garnett |
| Konstantinos Blekas | Jeroen De Knijf | Paolo Garza |
| Francesco Bonchi | Colin de la Higuera | Eric Gaussier |
| Gianluca Bontempi | Gerard de Melo | Floris Geerts |
| Christian Borgelt | Juan del Coz | Pierre Geurts |
| Marco Botta | Krzysztof Dembczyński | Rayid Ghani |
| Jean-François Boulicaut | François Denis | Fosca Giannotti |
| Marc Boullé | Anne Denton | Aris Gkoulalas-Divanis |
| Kendrick Boyd | Mohamed Dermouche | Vibhav Gogate |
| Pavel Brazdil | Christian Desrosiers | Marco Gori |
| Ulf Brefeld | Luigi Di Caro | Michael Granitzer |
| Björn Bringmann | Jana Diesner | Oded Green |

| | | |
|-----------------------|----------------------|------------------------|
| Tias Guns | Nicolas Lachiche | Benjamin Nguyen |
| Maria Halkidi | Pedro Larranaga | Thomas Niebler |
| Jiawei Han | Silvio Lattanzi | Thomas Nielsen |
| Daniel Hernandez | Niklas Lavesson | Siegfried Nijssen |
| Lobato | Nada Lavrač | Xia Ning |
| José Hernández-Orallo | Gregor Leban | Richard Nock |
| Frank Hoepfner | Sangkyun Lee | Niklas Noren |
| Jaakko Hollmén | Wang Lee | Kjetil Nørvåg |
| Geoff Holmes | Carson Leung | Eirini Ntoutsi |
| Arjen Hommersom | Jiuyong Li | Andreas Nürnberger |
| Vasant Honavar | Lei Li | Salvatore Orlando |
| Xiaohua Hu | Tao Li | Gerhard Paass |
| Minlie Huang | Rui Li | George Paliouras |
| Eyke Hüllermeier | Ping Li | Spiros Papadimitriou |
| Dino Ienco | Juanzi Li | Apostolos Papadopoulos |
| Robert Jäschke | Lei Li | Panagiotis Papapetrou |
| Frederik Janssen | Edo Liberty | Stelios Pappas |
| Nathalie Japkowicz | Jefrey Lijffijt | Ioannis Partalas |
| Szymon Jaroszewicz | shou-de Lin | Andrea Passerini |
| Ulf Johansson | Jessica Lin | Vladimir Pavlovic |
| Alipio Jorge | Hsuan-Tien Lin | Mykola Pechenizkiy |
| Kshitij Judah | Francesca Lisi | Dino Pedreschi |
| Tobias Jung | Yan Liu | Nikos Pelekis |
| Hachem Kadri | Huan Liu | Jing Peng |
| Theodore Kalamboukis | Corrado Loglisci | Ruggero Pensa |
| Alexandros Kalousis | Eneldo Loza Mencia | Fabio Pinelli |
| Pallika Kanani | Chang-Tien Lu | Marc Plantevit |
| U Kang | Panagis Magdalinos | Pascal Poncelet |
| Panagiotis Karras | Giuseppe Manco | George Potamias |
| Andreas Karwath | Yannis Manolopoulos | Aditya Prakash |
| Hisashi Kashima | Enrique Martinez | Doina Precup |
| Ioannis Katakis | Dimitrios Mavroeidis | Kai Puolamaki |
| John Keane | Mike Mayo | Buyue Qian |
| Latifur Khan | Wannes Meert | Chedy Raïssi |
| Levente Kocsis | Gabor Melli | Liva Ralaivola |
| Yun Sing Koh | Ernestina Menasalvas | Karthik Raman |
| Alek Kolcz | Roser Morante | Jan Ramon |
| Igor Kononenko | João Moreira | Huzefa Rangwala |
| Irena Koprinska | Emmanuel Müller | Zbigniew Raś |
| Nitish Korula | Mohamed Nadif | Chotirat |
| Petr Kosina | Mirco Nanni | Ratanamahatana |
| Walter Kosters | Alex Nanopoulos | Jan Rauch |
| Georg Kreml | Balakrishnan | Soumya Ray |
| Konstantin Kutzkov | Narayanaswamy | Steffen Rendle |
| Sergei Kuznetsov | Sriraam Natarajan | Achim Rettinger |

| | | |
|-----------------------|-----------------------|----------------------|
| Fabrizio Riguzzi | Johan Suykens | Fei Wang |
| Céline Robardet | Einoshin Suzuki | Jianyong Wang |
| Marko Robnik Sikonja | Panagiotis Symeonidis | Xiang Wang |
| Pedro Rodrigues | Sandor Szedmak | Yang Wang |
| Juan Rodriguez | Andrea Tagarelli | Takashi Washio |
| Irene Rodriguez-Lujan | Domenico Talia | Geoff Webb |
| Fabrice Rossi | Pang Tan | Jörg Wicker |
| Juho Rousu | Letizia Tanca | Hui Xiong |
| Céline Rouveirol | Dacheng Tao | Jieping Ye |
| Stefan Rüping | Nikolaž Tatti | Jeffrey Yu |
| Salvatore Ruggieri | Maguelonne Teisseire | Philip Yu |
| Yvan Saeys | Evimaria Terzi | Chun-Nam Yu |
| Alan Said | Martin Theobald | Jure Zabkar |
| Lorenza Saitta | Jilei Tian | Bianca Zadrozny |
| Ansaf Salleb-Aouissi | Ljupco Todorovski | Gerson Zaverucha |
| Scott Sanner | Luis Torgo | Demetris Zeinalipour |
| Vítor Santos Costa | Vicenç Torra | Filip Železný |
| Raul Santos-Rodriguez | Ivor Tsang | Bernard Zenko |
| Sam Sarjant | Panagiotis Tsaparas | Min-Ling Zhang |
| Claudio Sartori | Vincent Tseng | Nan Zhang |
| Taisuke Sato | Grigorios Tsoumakas | Zhongfei Zhang |
| Lars Schmidt-Thieme | Theodoros Tzouramanis | Junping Zhang |
| Christoph Schommer | Antti Ukkonen | Lei Zhang |
| Matthias Schubert | Takeaki Uno | Changshui Zhang |
| Giovanni Semeraro | Athina Vakali | Kai Zhang |
| Junming Shao | Giorgio Valentini | Kun Zhang |
| Junming Shao | Guy Van den Broeck | Shichao Zhang |
| Pannaga Shivaswamy | Peter van der Putten | Ying Zhao |
| Andrzej Skowron | Matthijs van Leeuwen | Elena Zheleva |
| Kevin Small | Maarten van Someren | Zhi-Hua Zhou |
| Padhraic Smyth | Joaquin Vanschoren | Bin Zhou |
| Carlos Soares | Iraklis Varlamis | Xingquan Zhu |
| Yangqiu Song | Michalis Vazirgiannis | Xiaofeng Zhu |
| Mauro Sozio | Julien Velcin | Kenny Zhu |
| Alessandro Sperduti | Shankar Vembu | Djamel Zighed |
| Eirini Spyropoulou | Sicco Verwer | Arthur Zimek |
| Jerzy Stefanowski | Vassilios Verykios | Albrecht Zimmermann |
| Jean Steyaert | Herna Viktor | Indre Zliobaite |
| Daniela Stojanova | Christel Vrain | Blaz Zupan |
| Markus Strohmaier | Willem Waegeman | |
| Mahito Sugiyama | Byron Wallace | |

Demo Track Program Committee

| | | |
|-------------------|----------------------|---------------------|
| Martin Atzmueller | Jaakko Hollmén | Mykola Pechenizkiy |
| Bettina Berendt | Andreas Hotho | Bernhard Pfahringer |
| Albert Bifet | Mark Last | Pedro Rodrigues |
| Antoine Bordes | Vincent Lemaire | Jerzy Stefanowski |
| Christian Borgelt | Ernestina Menasalvas | Grigorios Tsoumakas |
| Ulf Brefeld | Kjetil Nørvåg | Alice Zheng |
| Blaz Fortuna | Themis Palpanas | |

Nectar Track Program Committee

| | | |
|----------------|----------------|-------------------|
| Donato Malerba | George Karypis | Rosa Meo |
| Dora Erdos | Louis Wehenkel | Myra Spiliopoulou |
| Yiannis Koutis | Leman Akoglu | Toon Calders |

Additional Reviewers

| | | |
|----------------------|----------------------|------------------------|
| Argimiro Arratia | Elad Liebman | Kiumars Soltani |
| Rossella Cancelliere | Babak Loni | Ricardo Sousa |
| Antonio Corral | Emmanouil Magkos | Eleftherios |
| Joana Côte-Real | Adolfo Martínez-Usó | Spyromitros-Xioufis |
| Giso Dal | Dimitrios Mavroeidis | Jiliang Tang |
| Giacomo Domeniconi | Steffen Michels | Eleftherios Tiakas |
| Roberto Esposito | Pasquale Minervini | Andrei Tolstikov email |
| Pedro Ferreira | Fatemeh Mirrashed | Tiago Vinhoza |
| Asmelash Teka Hadgu | Fred Morstatter | Xing Wang |
| Isaac Jones | Tsuyoshi Murata | Lorenz Weizsäcker |
| Dimitris Kalles | Jinseok Nam | Sean Wilner |
| Yoshitaka Kameya | Rasaq Otunba | Christian Wirth |
| Eamonn Keogh | Roberto Pasolini | Lin Wu |
| Kristian Kersting | Tommaso Pirini | Jinfeng Yi |
| Rohan Khade | Maria-Jose | Cangzhou Yuan |
| Shamanth Kumar | Ramirez-Quintana | Jing Zhang |
| Hongfei Li | Irma Ravkic | |

Sponsors

Gold Sponsor

Winton

<http://www.wintoncapital.com>

Silver Sponsors

Deloitte <http://www.deloitte.com>
Xerox Research Centre Europe <http://www.xrce.xerox.com>

Bronze Sponsors

EDF <http://www.edf.com>
Orange <http://www.orange.com>
Technicolor <http://www.technicolor.com>
Yahoo! Labs <http://labs.yahoo.com>

Additional Supporters

Harmonic Pharma <http://www.harmonicpharma.com>
Deloitte <http://www.deloitte.com>

Lanyard

Knime <http://www.knime.org>

Prize

Deloitte <http://www.deloitte.com>
Data Mining and Knowledge
Discovery <http://link.springer.com/journal/10618>
Machine Learning <http://link.springer.com/journal/10994>

Organizing Institutions

Inria <http://www.inria.fr>
CNRS <http://www.cnrs.fr>
LORIA <http://www.loria.fr>

Invited Talks Abstracts

Scalable Collective Reasoning Using Probabilistic Soft Logic

Lise Getoor

University of California, Santa Cruz
Santa Cruz, CA, USA
getoor@cs.umd.edu

Abstract. One of the challenges in big data analytics is to efficiently learn and reason collectively about extremely large, heterogeneous, incomplete, noisy interlinked data. Collective reasoning requires the ability to exploit both the logical and relational structure in the data and the probabilistic dependencies. In this talk I will overview our recent work on probabilistic soft logic (PSL), a framework for collective, probabilistic reasoning in relational domains. PSL is able to reason holistically about both entity attributes and relationships among the entities. The underlying mathematical framework, which we refer to as a hinge-loss Markov random field, supports extremely efficient, exact inference. This family of graphical models captures logic-like dependencies with convex hinge-loss potentials. I will survey applications of PSL to diverse problems ranging from information extraction to computational social science. Our recent results show that by building on state-of-the-art optimization methods in a distributed implementation, we can solve large-scale problems with millions of random variables orders of magnitude faster than existing approaches.

Bio. In 1995, Lise Getoor decided to return to school to get her PhD in Computer Science at Stanford University. She received a National Physical Sciences Consortium fellowship, which in addition to supporting her for six years, supported a summer internship at Xerox PARC, where she worked with Markus Fromherz and his group. Daphne Koller was her PhD advisor; in addition, she worked closely with Nir Friedman, and many other members of the DAGS group, including Avi Pfeffer, Mehran Sahami, Ben Taskar, Carlos Guestrin, Uri Lerner, Ron Parr, Eran Segal, Simon Tong.

In 2001, Lise Getoor joined the Computer Science Department at the University of Maryland, College Park.

Network Analysis in the Big Data Age: Mining Graph and Social Streams

Charu Aggarwal

IBM T.J. Watson Research Center, New York
Yorktown, NY, USA
charu@us.ibm.com

Abstract. The advent of large interaction-based communication and social networks has led to challenging streaming scenarios in graph and social stream analysis. The graphs that result from such interactions are large, transient, and very often cannot even be stored on disk. In such cases, even simple frequency-based aggregation operations become challenging, whereas traditional mining operations are far more complex. When the graph cannot be explicitly stored on disk, mining algorithms must work with a limited knowledge of the network structure. Social streams add yet another layer of complexity, wherein the streaming content associated with the nodes and edges needs to be incorporated into the mining process. A significant gap exists between the problems that need to be solved, and the techniques that are available for streaming graph analysis. In spite of these challenges, recent years have seen some advances in which carefully chosen synopses of the graph and social streams are leveraged for approximate analysis. This talk will focus on several recent advances in this direction.

Bio. Charu Aggarwal is a Research Scientist at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his B.S. from IIT Kanpur in 1993 and his Ph.D. from Massachusetts Institute of Technology in 1996. His research interest during his Ph.D. years was in combinatorial optimization (network flow algorithms), and his thesis advisor was Professor James B. Orlin. He has since worked in the field of data mining, with particular interests in data streams, privacy, uncertain data and social network analysis. He has published over 200 papers in refereed venues, and has applied for or been granted over 80 patents. Because of the commercial value of the above-mentioned patents, he has received several invention achievement awards and has thrice been designated a Master Inventor at IBM. He is a recipient of an IBM Corporate Award (2003) for his work on bio-terrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and a recipient of an IBM Research Division Award (2008) for his scientific contributions to data stream research. He has served on the program committees of most major database/data mining conferences, and served as program vice-chairs of the SIAM Conference on Data Mining, 2007, the IEEE ICDM Conference, 2007, the WWW Conference 2009, and the IEEE ICDM Conference, 2009. He served as an associate editor of the IEEE Transactions on Knowledge

and Data Engineering Journal from 2004 to 2008. He is an associate editor of the ACM TKDD Journal, an action editor of the Data Mining and Knowledge Discovery Journal, an associate editor of the ACM SIGKDD Explorations, and an associate editor of the Knowledge and Information Systems Journal. He is a fellow of the ACM (2013) and the IEEE (2010) for contributions to knowledge discovery and data mining techniques.

Big Data for Personalized Medicine: A Case Study of Biomarker Discovery

Raymond Ng

University of British Columbia
Vancouver, B.C., Canada
mg@cs.ubc.ca

Abstract. Personalized medicine has been hailed as one of the main frontiers for medical research in this century. In the first half of the talk, we will give an overview on our projects that use gene expression, proteomics, DNA and clinical features for biomarker discovery. In the second half of the talk, we will describe some of the challenges involved in biomarker discovery. One of the challenges is the lack of quality assessment tools for data generated by ever-evolving genomics platforms. We will conclude the talk by giving an overview of some of the techniques we have developed on data cleansing and pre-processing.

Bio. Dr. Raymond Ng is a professor in Computer Science at the University of British Columbia. His main research area for the past two decades is on data mining, with a specific focus on health informatics and text mining. He has published over 180 peer-reviewed publications on data clustering, outlier detection, OLAP processing, health informatics and text mining. He is the recipient of two best paper awards from 2001 ACM SIGKDD conference, which is the premier data mining conference worldwide, and the 2005 ACM SIGMOD conference, which is one of the top database conferences worldwide. He was one of the program co-chairs of the 2009 International conference on Data Engineering, and one of the program co-chairs of the 2002 ACM SIGKDD conference. He was also one of the general co-chairs of the 2008 ACM SIGMOD conference. For the past decade, Dr. Ng has co-led several large scale genomic projects, funded by Genome Canada, Genome BC and industrial collaborators. The total amount of funding of those projects well exceeded \$40 million Canadian dollars. He now holds the Chief Informatics Officer position of the PROOF Centre of Excellence, which focuses on biomarker development for end-stage organ failures.

Machine Learning for Search Ranking and Ad Auction

Tie-Yan Liu

Microsoft Research Asia
Beijing, P.R. China
tyliu@microsoft.com

Abstract. In the era of information explosion, search has become an important tool for people to retrieve useful information. Every day, billions of search queries are submitted to commercial search engines. In response to a query, search engines return a list of relevant documents according to a ranking model. In addition, they also return some ads to users, and extract revenue by running an auction among advertisers if users click on these ads. This “search + ads” paradigm has become a key business model in today’s Internet industry, and has incubated a few hundred-billion-dollar companies. Recently, machine learning has been widely adopted in search and advertising, mainly due to the availability of huge amount of interaction data between users, advertisers, and search engines. In this talk, we discuss how to use machine learning to build effective ranking models (which we call learning to rank) and to optimize auction mechanisms. (i) The difficulty of learning to rank lies in the interdependency between documents in the ranked list. To tackle it, we propose the so-called listwise ranking algorithms, whose loss functions are defined on the permutations of documents, instead of individual documents or document pairs. We prove the effectiveness of these algorithms by analyzing their generalization ability and statistical consistency, based on the assumption of a two-layer probabilistic sampling procedure for queries and documents, and the characterization of the relationship between their loss functions and the evaluation measures used by search engines (e.g., NDCG and MAP). (ii) The difficulty of learning the optimal auction mechanism lies in that advertisers’ behavior data are strategically generated in response to the auction mechanism, but not randomly sampled in an i.i.d. manner. To tackle this challenge, we propose a game-theoretic learning method, which first models the strategic behaviors of advertisers, and then optimizes the auction mechanism by assuming the advertisers to respond to new auction mechanisms according to the learned behavior model. We prove the effectiveness of the proposed method by analyzing the generalization bounds for both behavior learning and auction mechanism learning based on a novel Markov framework.

Bio. Tie-Yan Liu is a senior researcher and research manager at Microsoft Research. His research interests include machine learning (learning to rank, online learning, statistical learning theory, and deep learning), algorithmic game theory, and computational economics. He is well known for his work on learning to rank

for information retrieval. He has authored the first book in this area, and published tens of highly-cited papers on both algorithms and theorems of learning to rank. He has also published extensively on other related topics. In particular, his paper won the best student paper award of SIGIR (2008), and the most cited paper award of the Journal of Visual Communication and Image Representation (2004-2006); his group won the research break-through award of Microsoft Research Asia (2012). Tie-Yan is very active in serving the research community. He is a program committee co-chair of ACML (2015), WINE (2014), AIRS (2013), and RIAO (2010), a local co-chair of ICML 2014, a tutorial co-chair of WWW 2014, a demo/exhibit co-chair of KDD (2012), and an area/track chair of many conferences including ACML (2014), SIGIR (2008-2011), AIRS (2009-2011), and WWW (2011). He is an associate editor of ACM Transactions on Information System (TOIS), an editorial board member of Information Retrieval Journal and Foundations and Trends in Information Retrieval. He has given keynote speeches at CCML (2013), CCIR (2011), and PCM (2010), and tutorials at SIGIR (2008, 2010, 2012), WWW (2008, 2009, 2011), and KDD (2012). He is a senior member of the IEEE and the ACM.

Beyond Stochastic Gradient Descent for Large-Scale Machine Learning

Francis Bach

INRIA, Paris
Laboratoire d'Informatique de l'Ecole Normale Supérieure
Paris, France
`francis.bach@inria.fr`

Abstract. Many machine learning and signal processing problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations (“large n ”) and each of these is large (“large p ”). In this setting, online algorithms such as stochastic gradient descent which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. In this talk, I will show how the smoothness of loss functions may be used to design novel algorithms with improved behavior, both in theory and practice: in the ideal infinite-data setting, an efficient novel Newton-based stochastic approximation algorithm leads to a convergence rate of $O(1/n)$ without strong convexity assumptions, while in the practical finite-data setting, an appropriate combination of batch and online algorithms leads to unexpected behaviors, such as a linear convergence rate for strongly convex problems, with an iteration cost similar to stochastic gradient descent.

(joint work with Nicolas Le Roux, Eric Moulines and Mark Schmidt)

Bio. Francis Bach is a researcher at INRIA, leading since 2011 the SIERRA project-team, which is part of the Computer Science Laboratory at Ecole Normale Supérieure. He completed his Ph.D. in Computer Science at U.C. Berkeley, working with Professor Michael Jordan, and spent two years in the Mathematical Morphology group at Ecole des Mines de Paris, then he joined the WILLOW project-team at INRIA/Ecole Normale Supérieure from 2007 to 2010. Francis Bach is interested in statistical machine learning, and especially in graphical models, sparse methods, kernel-based learning, convex optimization vision and signal processing.

Industrial Invited Talks Abstracts

Making Smart Metering Smarter by Applying Data Analytics

Georges Hébrail

EDF Lab
CLAMART, France
`georges.hebrail@edf.fr`

Abstract. New data is being collected from electric smart meters which are deployed in many countries. Electric power meters measure and transmit to a central information system electric power consumption from every individual household or enterprise. The sampling rate may vary from 10 minutes to 24 hours and the latency to reach the central information system may vary from a few minutes to 24h. This generates a large amount of - possibly streaming - data if we consider customers from an entire country (ex. 35 millions in France). This data is collected firstly for billing purposes but can be processed with data analytics tools with several other goals. The first part of the talk will recall the structure of electric power smart metering data and review the different applications which are considered today for applying data analytics to such data. In a second part of the talk, we will focus on a specific problem: spatio-temporal estimation of aggregated electric power consumption from incomplete metering data.

Bio. Georges Hébrail is a senior researcher at EDF Lab, the research centre of Electricité de France, one of the world's leading electric utility. His background is in Business Intelligence covering many aspects from data storage and querying to data analytics. From 2002 to 2010, he was a professor of computer science at Telecom ParisTech, teaching and doing research in the field of information systems and business intelligence, with a focus on time series management, stream processing and mining. His current research interest is on distributed and privacy-preserving data mining on electric power related data.

Ads That Matter

Alexandre Cotarmanac'h

VP Platform & Distribution

Twenga

alexandre.cotarmanach@twenga.com

Abstract. The advent of realtime bidding and online ad-exchanges has created a new and fast-growing competitive marketplace. In this new setting, media-buyers can make fine-grained decisions for each of the impressions being auctioned taking into account information from the context, the user and his/her past behavior. This new landscape is particularly interesting for online e-commerce players where user actions can also be measured online and thus allow for a complete measure of return on ad-spend.

Despite those benefits, new challenges need to be addressed such as:

- the design of a real-time bidding architecture handling high volumes of queries at low latencies,
- the exploration of a sparse and volatile high-dimensional space,
- as well as several statistical modeling problems (e.g. pricing, offer and creative selection).

In this talk, I will present an approach to realtime media buying for online e-commerce from our experience working in the field. I will review the aforementioned challenges and discuss open problems for serving ads that matter.

Bio. Alexandre Cotarmanac'h is Vice-President Distribution & Platform for Twenga.

Twenga is a services and solutions provider generating high value-added leads to online merchants that was founded in 2006.

Originally hired to help launch Twenga's second generation search engine and to manage the optimization of revenue, he launched in 2011 the affinitAD line of business and Twenga's publisher network. Thanks to the advanced contextual analysis which allows for targeting the right audience according to their desire to buy e-commerce goods whilst keeping in line with the content offered, affinitAD brings Twenga's e-commerce expertise to web publishers. Alexandre also oversees Twenga's merchant programme and strives to offer Twenga's merchants new services and solutions to improve their acquisition of customers.

With over 14 years of experience, Alexandre has held a succession of increasingly responsible positions focusing on advertising and web development. Prior to joining Twenga, he was responsible for the development of Search and Advertising at Orange. Alexandre graduated from Ecole polytechnique.

Machine Learning and Data Mining in Call of Duty

Arthur Von Eschen

Activision Publishing Inc.
Santa Monica, CA, USA

Arthur.VonEschen@activision.com

Abstract. Data science is relatively new to the video game industry, but it has quickly emerged as one of the main resources for ensuring game quality. At Activision, we leverage data science to analyze the behavior of our games and our players to improve in-game algorithms and the player experience. We use machine learning and data mining techniques to influence creative decisions and help inform the game design process. We also build analytic services that support the game in real-time; one example is a cheating detection system which is very similar to fraud detection systems used for credit cards and insurance. This talk will focus on our data science work for Call of Duty, one of the bestselling video games in the world.

Bio. Arthur Von Eschen is Senior Director of Game Analytics at Activision. He and his team are responsible for analytics work that supports video game design on franchises such as Call of Duty and Skylanders. In addition to holding a PhD in Operations Research, Arthur has over 15 years of experience in analytics consulting and R&D with the U.S. Fortune 500. His work has spanned across industries such as banking, financial services, insurance, retail, CPG and now interactive entertainment (video games). Prior to Activision he worked at Fair Isaac Corporation (FICO). Before FICO he ran his own analytics consulting firm for six years.

Algorithms, Evolution and Network-Based Approaches in Molecular Discovery

Mike Bodkin

Evotec Ltd.
Oxfordshire, UK

Mike.Bodkin@evotec.com

Abstract. Drug research generates huge quantities of data around targets, compounds and their effects. Network modelling can be used to describe such relationships with the aim to couple our understanding of disease networks with the changes in small molecule properties. This talk will build off of the data that is routinely captured in drug discovery and describe the methods and tools that we have developed for compound design using predictive modelling, evolutionary algorithms and network-based mining.

Bio. Mike did his PhD in protein de-novo design for Nobel laureate sir James Black before taking up a fellowship in computational drug design at Cambridge University. He moved to AstraZeneca as a computational chemist before joining Eli Lilly in 2000. As head of the computational drug discovery group at Lilly since 2003 he recently jumped ship to Evotec to work as the VP for computational chemistry and cheminformatics. His research aims are to continue to develop new algorithms and software in the fields of drug discovery and systems informatics and to deliver and apply current and novel methods as tools for use in drug research.

Table of Contents – Part I

Main Track Contributions

| | |
|--|-----|
| Classifying a Stream of Infinite Concepts: A Bayesian Non-parametric Approach | 1 |
| <i>Seyyed Abbas Hosseini, Hamid R. Rabiee, Hassan Hafez, and Ali Soltani-Farani</i> | |
| Fast Nearest Neighbor Search on Large Time-Evolving Graphs | 17 |
| <i>Leman Akoglu, Rohit Khandekar, Vibhore Kumar, Srinivasan Parthasarathy, Deepak Rajan, and Kun-Lung Wu</i> | |
| Deconstructing Kernel Machines | 34 |
| <i>Mohsen Ali, Muhammad Rushdi, and Jeffrey Ho</i> | |
| Beyond Blocks: Hyperbolic Community Detection | 50 |
| <i>Miguel Araujo, Stephan Günnemann, Gonzalo Mateos, and Christos Faloutsos</i> | |
| Concurrent Bandits and Cognitive Radio Networks | 66 |
| <i>Orly Avner and Shie Mannor</i> | |
| Kernel Principal Geodesic Analysis | 82 |
| <i>Suyash P. Awate, Yen-Yun Yu, and Ross T. Whitaker</i> | |
| Attributed Graph Kernels Using the Jensen-Tsallis q -Differences | 99 |
| <i>Lu Bai, Luca Rossi, Horst Bunke, and Edwin R. Hancock</i> | |
| Sub-sampling for Multi-armed Bandits | 115 |
| <i>Akram Baransi, Odalric-Ambrym Maillard, and Shie Mannor</i> | |
| Knowledge-Powered Deep Learning for Word Embedding | 132 |
| <i>Jiang Bian, Bin Gao, and Tie-Yan Liu</i> | |
| Density-Based Subspace Clustering in Heterogeneous Networks | 149 |
| <i>Brigitte Boden, Martin Ester, and Thomas Seidl</i> | |
| Open Question Answering with Weakly Supervised Embedding Models | 165 |
| <i>Antoine Bordes, Jason Weston, and Nicolas Usunier</i> | |
| Towards Automatic Feature Construction for Supervised Classification | 181 |
| <i>Marc Boullé</i> | |

| | |
|--|-----|
| Consistency of Losses for Learning from Weak Labels | 197 |
| <i>Jesús Cid-Sueiro, Darío García-García, and Raúl Santos-Rodríguez</i> | |
| Active Learning for Support Vector Machines with Maximum Model Change | 211 |
| <i>Wenbin Cai, Ya Zhang, Siyuan Zhou, Wenquan Wang, Chris Ding, and Xiao Gu</i> | |
| Anomaly Detection with Score Functions Based on the Reconstruction Error of the Kernel PCA | 227 |
| <i>Laetitia Chapel and Chloé Friguet</i> | |
| Fast Gaussian Pairwise Constrained Spectral Clustering | 242 |
| <i>David Chatel, Pascal Denis, and Marc Tommasi</i> | |
| A Generative Bayesian Model for Item and User Recommendation in Social Rating Networks with Trust Relationships | 258 |
| <i>Gianni Costa, Giuseppe Manco, and Riccardo Ortale</i> | |
| Domain Adaptation with Regularized Optimal Transport | 274 |
| <i>Nicolas Courty, Rémi Flamary, and Devis Tuia</i> | |
| Discriminative Subnetworks with Regularized Spectral Learning for Global-State Network Data | 290 |
| <i>Xuan Hong Dang, Ambuj K. Singh, Petko Bogdanov, Hongyuan You, and Bayyuan Hsu</i> | |
| Infinitely Many-Armed Bandits with Unknown Value Distribution | 307 |
| <i>Yahel David and Nahum Shimkin</i> | |
| Cautious Ordinal Classification by Binary Decomposition | 323 |
| <i>Sébastien Destercke and Gen Yang</i> | |
| Error-Bounded Approximations for Infinite-Horizon Discounted Decentralized POMDPs | 338 |
| <i>Jilles S. Dibangoye, Olivier Buffet, and François Charpillet</i> | |
| Approximate Consistency: Towards Foundations of Approximate Kernel Selection | 354 |
| <i>Lizhong Ding and Shizhong Liao</i> | |
| Nowcasting with Numerous Candidate Predictors | 370 |
| <i>Brendan Duncan and Charles Elkan</i> | |
| Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries | 386 |
| <i>Flavio Figueiredo, Jussara M. Almeida, Yasuko Matsubara, Bruno Ribeiro, and Christos Faloutsos</i> | |

| | |
|--|-----|
| FASOLE: Fast Algorithm for Structured Output LEarning | 402 |
| <i>Vojtech Franc</i> | |
| Neutralized Empirical Risk Minimization with Generalization | |
| Neutrality Bound | 418 |
| <i>Kazuto Fukuchi and Jun Sakuma</i> | |
| Effective Blending of Two and Three-way Interactions for Modeling | |
| Multi-relational Data | 434 |
| <i>Alberto García-Durán, Antoine Bordes, and Nicolas Usunier</i> | |
| Automatic Design of Neuromarkers for OCD Characterization | 450 |
| <i>Oscar García Hinde, Emilio Parrado-Hernández, Vanessa Gómez-Verdejo, Manel Martínez-Ramón, and Carles Soriano-Mas</i> | |
| Importance Weighted Inductive Transfer Learning for Regression | 466 |
| <i>Jochen Garcke and Thomas Vanck</i> | |
| Policy Search for Path Integral Control | 482 |
| <i>Vicenç Gómez, Hilbert J. Kappen, Jan Peters, and Gerhard Neumann</i> | |
| How Many Topics? Stability Analysis for Topic Models | 498 |
| <i>Derek Greene, Derek O’Callaghan, and Pádraig Cunningham</i> | |
| Joint Prediction of Topics in a URL Hierarchy | 514 |
| <i>Michael Großhans, Christoph Sawade, Tobias Scheffer, and Niels Landwehr</i> | |
| Learned-Norm Pooling for Deep Feedforward and Recurrent Neural | |
| Networks | 530 |
| <i>Caglar Gulcehre, Kyunghyun Cho, Razvan Pascanu, and Yoshua Bengio</i> | |
| Combination of One-Class Support Vector Machines for Classification | |
| with Reject Option | 547 |
| <i>Blaise Hanczar and Michèle Sebag</i> | |
| Relative Comparison Kernel Learning with Auxiliary Kernels | 563 |
| <i>Eric Heim, Hamed Valizadegan, and Milos Hauskrecht</i> | |
| Transductive Minimax Probability Machine | 579 |
| <i>Gao Huang, Shiji Song, Zhixiang (Eddie) Xu, and Kilian Weinberger</i> | |
| Covariate-Related Lasso for Feature Selection | 595 |
| <i>Bo Jiang, Chris Ding, and Bin Luo</i> | |
| Random Forests with Random Projections of the Output Space for | |
| High Dimensional Multi-label Classification | 607 |
| <i>Arnaud Joly, Pierre Geurts, and Louis Wehenkel</i> | |

| | |
|--|-----|
| Communication-Efficient Distributed Online Prediction by Dynamic Model Synchronization | 623 |
| <i>Michael Kamp, Mario Boley, Daniel Keren, Assaf Schuster, and Izhak Sharfman</i> | |
| Hetero-Labeled LDA: A Partially Supervised Topic Model with Heterogeneous Labels | 640 |
| <i>Dongyeop Kang, Youngja Park, and Suresh N. Chari</i> | |
| Bayesian Multi-view Tensor Factorization | 656 |
| <i>Suleiman A. Khan and Samuel Kaski</i> | |
| Conditional Log-Linear Models for Mobile Application Usage Prediction | 672 |
| <i>Jingu Kim and Taneli Mielikäinen</i> | |
| Pushing-Down Tensor Decompositions over Unions to Promote Reuse of Materialized Decompositions | 688 |
| <i>Mijung Kim and K. Selçuk Candan</i> | |
| Author Index | 705 |

Table of Contents – Part II

Main Track Contributions

| | |
|--|-----|
| Robust Distributed Training of Linear Classifiers Based on Divergence Minimization Principle | 1 |
| <i>Junpei Komiyama, Hidekazu Oiwa, and Hiroshi Nakagawa</i> | |
| Reliability Maps: A Tool to Enhance Probability Estimates and Improve Classification Accuracy | 18 |
| <i>Meelis Kull and Peter A. Flach</i> | |
| Causal Clustering for 2-Factor Measurement Models | 34 |
| <i>Erich Kummerfeld, Joe Ramsey, Renjie Yang, Peter Spirtes, and Richard Scheines</i> | |
| Support Vector Machines for Differential Prediction | 50 |
| <i>Finn Kuusisto, Vitor Santos Costa, Houssam Nassif, Elizabeth Burnside, David Page, and Jude Shavlik</i> | |
| Fast LSTD Using Stochastic Approximation: Finite Time Analysis and Application to Traffic Control | 66 |
| <i>L.A. Prashanth, Nathaniel Korda, and Rémi Munos</i> | |
| Mining Top-K Largest Tiles in a Data Stream | 82 |
| <i>Hoang Thanh Lam, Wenjie Pei, Adriana Prado, Baptiste Jeudy, and Élisabeth Fromont</i> | |
| Ranked Tiling | 98 |
| <i>Thanh Le Van, Matthijs van Leeuwen, Siegfried Nijssen, Ana Carolina Fierro, Kathleen Marchal, and Luc De Raedt</i> | |
| Fast Estimation of the Pattern Frequency Spectrum | 114 |
| <i>Matthijs van Leeuwen and Antti Ukkonen</i> | |
| Recurrent Greedy Parsing with Neural Networks | 130 |
| <i>Joël Legrand and Ronan Collobert</i> | |
| FILTA: Better View Discovery from Collections of Clusterings via Filtering | 145 |
| <i>Yang Lei, Nguyen Xuan Vinh, Jeffrey Chan, and James Bailey</i> | |
| Nonparametric Markovian Learning of Triggering Kernels for Mutually Exciting and Mutually Inhibiting Multivariate Hawkes Processes | 161 |
| <i>Remi Lemonnier and Nicolas Vayatis</i> | |

| | |
|---|-----|
| Learning Binary Codes with Bagging PCA | 177 |
| <i>Cong Leng, Jian Cheng, Ting Yuan, Xiao Bai, and Hanqing Lu</i> | |
| Conic Multi-task Classification | 193 |
| <i>Cong Li, Michael Georgiopoulos, and Georgios C. Anagnostopoulos</i> | |
| Bi-directional Representation Learning for Multi-label Classification | 209 |
| <i>Xin Li and Yuhong Guo</i> | |
| Optimal Thresholding of Classifiers to Maximize F1 Measure | 225 |
| <i>Zachary C. Lipton, Charles Elkan, and Balakrishnan Naryanaswamy</i> | |
| Randomized Operating Point Selection in Adversarial Classification | 240 |
| <i>Viliam Lisý, Robert Kessl, and Tomáš Pevný</i> | |
| Hierarchical Latent Tree Analysis for Topic Detection | 256 |
| <i>Tengfei Liu, Nevin L. Zhang, and Peixian Chen</i> | |
| Bayesian Models for Structured Sparse Estimation via Set Cover Prior | 273 |
| <i>Xianghang Liu, Xinhua Zhang, and Tibério Caetano</i> | |
| Preventing Over-Fitting of Cross-Validation with Kernel Stability | 290 |
| <i>Yong Liu and Shizhong Liao</i> | |
| Experimental Design in Dynamical System Identification: A Bandit-Based Active Learning Approach | 306 |
| <i>Artémis Llamosi, Adel Mezine, Florence d'Alché-Buc, Véronique Letort, and Michèle Sebag</i> | |
| On the Null Distribution of the Precision and Recall Curve | 322 |
| <i>Miguel Lopes and Gianluca Bontempi</i> | |
| Linear State-Space Model with Time-Varying Dynamics | 338 |
| <i>Jaakko Luttinen, Tapani Raiko, and Alexander Ilin</i> | |
| An Online Policy Gradient Algorithm for Markov Decision Processes with Continuous States and Actions | 354 |
| <i>Yao Ma, Tingting Zhao, Kohei Hatano, and Masashi Sugiyama</i> | |
| GMRF Estimation under Topological and Spectral Constraints | 370 |
| <i>Victorin Martin, Cyril Furtlehner, Yufei Han, and Jean-Marc Lasgouttes</i> | |
| Rate-Constrained Ranking and the Rate-Weighted AUC | 386 |
| <i>Louise A.C. Millard, Peter A. Flach, and Julian P.T. Higgins</i> | |
| Rate-Oriented Point-Wise Confidence Bounds for ROC Curves | 404 |
| <i>Louise A.C. Millard, Meelis Kull, and Peter A. Flach</i> | |

| | |
|---|-----|
| A Fast Method of Statistical Assessment for Combinatorial Hypotheses Based on Frequent Itemset Enumeration | 422 |
| <i>Shin-ichi Minato, Takeaki Uno, Koji Tsuda, Aika Terada, and Jun Sese</i> | |
| Large-Scale Multi-label Text Classification — Revisiting Neural Networks | 437 |
| <i>Jinseok Nam, Jungi Kim, Eneldo Loza Mencía, Iryna Gurevych, and Johannes Fürnkranz</i> | |
| Distinct Chains for Different Instances: An Effective Strategy for Multi-label Classifier Chains | 453 |
| <i>Pablo Nascimento da Silva, Eduardo Corrêa Gonçalves, Alexandre Plastino, and Alex A. Freitas</i> | |
| A Unified Framework for Probabilistic Component Analysis | 469 |
| <i>Mihalis A. Nicolaou, Stefanos Zafeiriou, and Maja Pantic</i> | |
| Flexible Shift-Invariant Locality and Globality Preserving Projections | 485 |
| <i>Feiping Nie, Xiao Cai, and Heng Huang</i> | |
| Interactive Knowledge-Based Kernel PCA | 501 |
| <i>Dino Ogljic, Daniel Paurat, and Thomas Gärtner</i> | |
| A Two-Step Learning Approach for Solving Full and Almost Full Cold Start Problems in Dyadic Prediction | 517 |
| <i>Tapio Pahikkala, Michiel Stock, Antti Airola, Tero Aittokallio, Bernard De Baets, and Willem Waegeman</i> | |
| Deterministic Feature Selection for Regularized Least Squares Classification | 533 |
| <i>Saurabh Paul and Petros Drineas</i> | |
| Boosted Bellman Residual Minimization Handling Expert Demonstrations | 549 |
| <i>Bilal Piot, Matthieu Geist, and Olivier Pietquin</i> | |
| Semi-Supervised Learning Using an Unsupervised Atlas | 565 |
| <i>Nikolaos Pitelis, Chris Russell, and Lourdes Agapito</i> | |
| A Lossless Data Reduction for Mining Constrained Patterns in n -ary Relations | 581 |
| <i>Gabriel Poesia and Loïc Cerf</i> | |
| Interestingness-Driven Diffusion Process Summarization in Dynamic Networks | 597 |
| <i>Qiang Qu, Siyuan Liu, Christian S. Jensen, Feida Zhu, and Christos Faloutsos</i> | |

| | |
|---|-----|
| Neural Gaussian Conditional Random Fields | 614 |
| <i>Vladan Radosavljevic, Slobodan Vucetic, and Zoran Obradovic</i> | |
| Cutset Networks: A Simple, Tractable, and Scalable Approach for Improving the Accuracy of Chow-Liu Trees | 630 |
| <i>Tahrima Rahman, Prasanna Kothalkar, and Vibhav Gogate</i> | |
| Boosted Mean Shift Clustering | 646 |
| <i>Yazhou Ren, Uday Kamath, Carlotta Domeniconi, and Guoji Zhang</i> | |
| Hypernode Graphs for Spectral Learning on Binary Relations over Sets | 662 |
| <i>Thomas Ricatte, Rémi Gilleron, and Marc Tommasi</i> | |
| Discovering Dynamic Communities in Interaction Networks | 678 |
| <i>Polina Rozenshtein, Nikolaj Tatti, and Aristides Gionis</i> | |
| Anti-discrimination Analysis Using Privacy Attack Strategies | 694 |
| <i>Salvatore Ruggieri, Sara Hajian, Faisal Kamiran, and Xiangliang Zhang</i> | |
| Author Index | 711 |

Table of Contents – Part III

Main Track Contributions

| | |
|--|-----|
| FLIP: Active Learning for Relational Network Classification | 1 |
| <i>Tanwistha Saha, Huzefa Rangwala, and Carlotta Domeniconi</i> | |
| Clustering via Mode Seeking by Direct Estimation of the Gradient of a Log-Density | 19 |
| <i>Hiroaki Sasaki, Aapo Hyvärinen, and Masashi Sugiyama</i> | |
| Local Policy Search in a Convex Space and Conservative Policy Iteration as Boosted Policy Search | 35 |
| <i>Bruno Scherrer and Matthieu Geist</i> | |
| Code You Are Happy to Paste: An Algorithmic Dictionary of Exponential Families | 51 |
| <i>Olivier Schwander</i> | |
| Statistical Hypothesis Testing in Positive Unlabelled Data | 66 |
| <i>Konstantinos Sechidis, Borja Calvo, and Gavin Brown</i> | |
| Students, Teachers, Exams and MOOCs: Predicting and Optimizing Attainment in Web-Based Education Using a Probabilistic Graphical Model | 82 |
| <i>Bar Shalem, Yoram Bachrach, John Guiver, and Christopher M. Bishop</i> | |
| Gaussian Process Multi-task Learning Using Joint Feature Selection | 98 |
| <i>P.K. Srijith and Shirish Shevade</i> | |
| Separating Rule Refinement and Rule Selection Heuristics in Inductive Rule Learning | 114 |
| <i>Julius Stecher, Frederik Janssen, and Johannes Fürnkranz</i> | |
| Scalable Information Flow Mining in Networks | 130 |
| <i>Karthik Subbian, Chidananda Sridhar, Charu C. Aggarwal, and Jaideep Srivastava</i> | |
| Link Prediction in Multi-modal Social Networks | 147 |
| <i>Panagiotis Symeonidis and Christos Perentis</i> | |
| Faster Way to Agony: Discovering Hierarchies in Directed Graphs | 163 |
| <i>Nikolaj Tatti</i> | |

| | |
|---|-----|
| Speeding Up Recovery from Concept Drifts | 179 |
| <i>Silas Garrido Teixeira de Carvalho Santos,</i> | |
| <i>Paulo Mauricio Gonçalves Júnior,</i> | |
| <i>Geyson Daniel dos Santos Silva, and</i> | |
| <i>Roberto Souto Maior de Barros</i> | |
| Training Restricted Boltzmann Machines with Overlapping Partitions | 195 |
| <i>Hasari Tosun and John W. Sheppard</i> | |
| Integer Bayesian Network Classifiers | 209 |
| <i>Sebastian Tschiatschek, Karin Paul, and Franz Pernkopf</i> | |
| Multi-target Regression via Random Linear Target Combinations | 225 |
| <i>Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis,</i> | |
| <i>Aikaterini Vrekou, and Ioannis Vlahavas</i> | |
| Ratio-Based Multiple Kernel Clustering | 241 |
| <i>Grigorios Tzortzis and Aristidis Likas</i> | |
| Evidence-Based Clustering for Scalable Inference in Markov Logic | 258 |
| <i>Deepak Venugopal and Vibhav Gogate</i> | |
| On Learning Matrices with Orthogonal Columns or Disjoint Supports | 274 |
| <i>Kevin Vervier, Pierre Mahé, Alexandre D’Aspremont,</i> | |
| <i>Jean-Baptiste Veyrieras, and Jean-Philippe Vert</i> | |
| Scalable Moment-Based Inference for Latent Dirichlet Allocation | 290 |
| <i>Chi Wang, Xueqing Liu, Yanglei Song, and Jiawei Han</i> | |
| Unsupervised Feature Selection via Unified Trace Ratio Formulation and K -means Clustering (TRACK) | 306 |
| <i>De Wang, Feiping Nie, and Heng Huang</i> | |
| On the Equivalence between Deep NADE and Generative Stochastic Networks | 322 |
| <i>Li Yao, Sherjil Ozair, Kyunghyun Cho, and Yoshua Bengio</i> | |
| Scalable Nonnegative Matrix Factorization with Block-wise Updates | 337 |
| <i>Jiangtao Yin, Lixin Gao, and Zhongfei (Mark) Zhang</i> | |
| Convergence of Min-Sum-Min Message-Passing for Quadratic Optimization | 353 |
| <i>Guoqiang Zhang and Richard Heusdens</i> | |
| Clustering Image Search Results by Entity Disambiguation | 369 |
| <i>Kaiqi Zhao, Zhiyuan Cai, Qingyu Sui, Enxun Wei, and</i> | |
| <i>Kenny Q. Zhu</i> | |

| | |
|---|-----|
| Accelerating Model Selection with Safe Screening for L_1 -Regularized L_2 -SVM | 385 |
| <i>Zheng Zhao, Jun Liu, and James Cox</i> | |
| Kernel Alignment Inspired Linear Discriminant Analysis | 401 |
| <i>Shuai Zheng and Chris Ding</i> | |
| Transfer Learning with Multiple Sources via Consensus Regularized Autoencoders | 417 |
| <i>Fuzhen Zhuang, Xiaohu Cheng, Sinno Jialin Pan, Wenchao Yu, Qing He, and Zhongzhi Shi</i> | |

Demo Track Contributions

| | |
|---|-----|
| Branty: A Social Media Ranking Tool for Brands | 432 |
| <i>Alexandros Arvanitidis, Anna Serafi, Athena Vakali, and Grigorios Tsoumakas</i> | |
| MinUS: Mining User Similarity with Trajectory Patterns | 436 |
| <i>Xihui Chen, Piotr Kordy, Ruipeng Lu, and Jun Pang</i> | |
| KnowNow: A Serendipity-Based Educational Tool for Learning Time-Linked Knowledge | 440 |
| <i>Luigi Di Caro, Livio Robaldo, and Nicoletta Bersia</i> | |
| Khiops CoViz: A Tool for Visual Exploratory Analysis of k -Coclustering Results | 444 |
| <i>Bruno Guerraz, Marc Boullé, Dominique Gay, and Fabrice Clérot</i> | |
| PYTHIA: Employing Lexical and Semantic Features for Sentiment Analysis | 448 |
| <i>Ioannis Manoussos Katakis, Iraklis Varlamis, and George Tsatsaronis</i> | |
| Spá: A Web-Based Viewer for Text Mining in Evidence Based Medicine | 452 |
| <i>J. Kuiper, I.J. Marshall, B.C. Wallace, and M.A. Swertz</i> | |
| Propositionalization Online | 456 |
| <i>Nada Lavrač, Matic Perovšek, and Anže Vavpetič</i> | |
| Interactive Medical Miner: Interactively Exploring Subpopulations in Epidemiological Datasets | 460 |
| <i>Uli Niemann, Myra Spiliopoulou, Henry Völzke, and Jens-Peter Kühn</i> | |
| WebDR: A Web Workbench for Data Reduction | 464 |
| <i>Stefanos Ogiaroglou and Georgios Evangelidis</i> | |

GrammarViz 2.0: A Tool for Grammar-Based Pattern Discovery in Time Series 468
Pavel Senin, Jessica Lin, Xing Wang, Tim Oates, Sunil Gandhi, Arnold P. Boedihardjo, Crystal Chen, Susan Frankenstein, and Manfred Lerner

Insight4News: Connecting News to Relevant Social Conversations 473
Bichen Shi, Georgiana Ifrim, and Neil Hurley

BestTime: Finding Representatives in Time Series Datasets 477
Stephan Spiegel, David Schultz, and Sahin Albayrak

BMAD – A Boolean Matrix Decomposition Framework 481
Andrey Tyukin, Stefan Kramer, and Jörg Wicker

Nectar Track Contributions

Analyzing and Grounding Social Interaction in Online and Offline Networks 485
Martin Atzmueller

Be *Certain* of How-to before Mining *Uncertain* data 489
Francesco Gullo, Giovanni Ponti, and Andrea Tagarelli

Active Learning Is Planning: Nonmyopic ϵ -Bayes-Optimal Active Learning of Gaussian Processes 494
Trong Nghia Hoang, Kian Hsiang Low, Patrick Jaillet, and Mohan Kankanhalli

Generalized Online Sparse Gaussian Processes with Application to Persistent Mobile Robot Localization 499
Kian Hsiang Low, Nuo Xu, Jie Chen, Keng Kiat Lim, and Etkin Bariş Özgül

Distributional Clauses Particle Filter 504
Davide Nitti, Tinne De Laet, and Luc De Raedt

Network Reconstruction for the Identification of miRNA:mRNA Interaction Networks 508
Gianvito Pio, Michelangelo Ceci, Domenica D’Elia, and Donato Malerba

Machine Learning Approaches for Metagenomics 512
Huzefa Rangwala, Anveshi Charuvaka, and Zeehasham Rasheed

Sampling-Based Data Mining Algorithms: Modern Techniques and Case Studies 516
Matteo Riondato

| | |
|--|-----|
| Heterogeneous Stream Processing and Crowdsourcing for Traffic Monitoring: Highlights | 520 |
| <i>François Schnitzler, Alexander Artikis, Matthias Weidlich, Ioannis Boutsis, Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik, Vana Kalogeraki, Jakub Marecek, Avigdor Gal, Shie Mannor, Dermot Kinane, and Dimitrios Gunopulos</i> | |
| Agents Teaching Agents in Reinforcement Learning (Nectar Abstract) | 524 |
| <i>Matthew E. Taylor and Lisa Torrey</i> | |
| Author Index | 529 |