

Lecture Notes in Artificial Intelligence 5211

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Walter Daelemans Bart Goethals
Katharina Morik (Eds.)

Machine Learning and Knowledge Discovery in Databases

European Conference, ECML PKDD 2008
Antwerp, Belgium, September 15-19, 2008
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

Walter Daelemans
University of Antwerp, Department of Linguistics
CNTS Language Technology Group
Prinsstraat 13, L-203, 2000 Antwerp, Belgium
E-mail: walter.daelemans@ua.ac.be

Bart Goethals
University of Antwerp
Mathematics and Computer Science Department
Middelheimlaan 1, 2020 Antwerp, Belgium
E-mail: bart.goethals@ua.ac.be

Katharina Morik
Technische Universität Dortmund
Computer Science VIII, Artificial Intelligence Unit
44221 Dortmund, Germany
E-mail: katharina.morik@tu-dortmund.de

The copyright of the photo on the cover belongs to Tourism Antwerp.

Library of Congress Control Number: 2008934755

CR Subject Classification (1998): I.2, H.2.8, H.2, H.3, G.3, J.1, I.7, F.2.2, F.4.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-87478-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-87478-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12521968 06/3180 5 4 3 2 1 0

Preface

When in 1986 Yves Kodratoff started the European Working Session on Learning at Orsay, France, it could not be foreseen that the conference would grow year by year and become the premier European conference of the field, attracting submissions from all over the world. The first European Conference on Principles of Data Mining and Knowledge Discovery was organized by Henryk Jan Komorowski and Jan Zytkow in 1997 in Trondheim, Norway. Since 2001 the two conferences have been collocated, offering participants from both areas the opportunity to listen to each other's talks. This year, the integration has moved even further. Instead of first splitting the field according to ECML or PKDD topics, we flattened the structure of the field to a single set of topics. For each of the topics, experts were invited to the Program Committee. Submitted papers were gathered into one collection and characterized according to their topics. The reviewers were then asked to bid on all papers, regardless of the conference. This allowed us to allocate papers more precisely.

The hierarchical reviewing process as introduced in 2005 was continued. We nominated 30 Area Chairs, each supervising the reviews and discussions of about 17 papers. In addition, 307 reviewers completed the Program Committee. Many thanks to all of them! It was a considerable effort for the reviewers to carefully review the papers, some providing us with additional reviews even at short notice. Based on their reviews and internal discussions, which were concluded by the recommendations of the Area Chairs, we could manage the final selection for the program. We received 521 submissions, of which 100 were presented at the conferences, giving us an acceptance rate of 20%. This high selectivity means, on the one hand, that some good papers could not make it into the conference program. On the other hand, it supports the traditionally high standards of the joint conference. We thank the authors from all over the world for submitting their contributions!

Following the tradition, the first and the last day of the joint conference were dedicated to workshops and tutorials. ECML PKDD 2008 offered 8 tutorials and 11 workshops. We thank the Workshop and Tutorial Chairs Siegfried Nijssen and Arno Siebes for their excellent selection. The discovery challenge is also a tradition of ECML PKDD that we continued. We are grateful to Andreas Hotho and his colleagues from the Bibsonomy project for organizing the discovery challenge of this year. The results were presented at the Web 2.0 Mining Workshop.

One of the pleasures of chairing a conference is the opportunity to invite colleagues whose work we esteem highly. We are grateful to Françoise Fogelman Soulié (KXEN) for opening the industrial track, Yoav Freund (University of California, San Diego), Anil K. Jain (Michigan State University), Ray Mooney (University of Texas at Austin), and Raghu Ramakrishnan (Yahoo! Research) for accepting our invitation to present recent work at the conference.

Some novelties were introduced to the joint conference this year.

First, there was no distinction into long and short papers. Instead, paper length was raised to 16 pages for all submissions.

Second, 14 papers were selected for publication in Springer Journals. Seven papers were published in the *Machine Learning Journal* 72:3 (September 2008), and 7 papers were published in the *Data Mining and Knowledge Discovery Journal* 17:1 (August 2008). This LNAI volume includes the abstracts of these papers, each containing a reference to the respective full journal contribution. At the conference, participants received the proceedings, the tutorial notes and workshop proceedings on a USB memory stick.

Third, all papers were additionally allowed to be presented as posters. Since the number of participants has become larger, questions and discussions after a talk are no longer possible for all those interested. Introducing poster presentations for all accepted papers allows for more detailed discussions. Hence, we did not reserve this opportunity for a minority of papers and it was not an alternative to an oral presentation.

Fourth, a special demonstration session was held that is intended to be a forum for showcasing the state of the art in machine learning and knowledge discovery software. The focus lies on innovative prototype implementations in machine learning and data analysis. The demo descriptions are included in the proceedings. We thank Christian Borgelt for reviewing the submitted demos. Finally, for the first time, the conference took place in Belgium!

September 2008

Walter Daelemans
Bart Goethals
Katharina Morik

Organization

Program Chairs

Walter Daelemans	University of Antwerp, Belgium
Bart Goethals	University of Antwerp, Belgium
Katharina Morik	Technische Universität Dortmund, Germany

Workshop and Tutorial Chairs

Siegfried Nijssen	Utrecht University, the Netherlands
Arno Siebes	Katholieke Universiteit Leuven, Belgium

Demo Chair

Christian Borgelt	European Centre for Soft Computing, Spain
-------------------	---

Publicity Chairs

Hendrik Blockeel	Katholieke Universiteit Leuven, Belgium
Pierre Dupont	Université catholique de Louvain, Belgium
Jef Wijsen	University of Mons-Hainaut, Belgium

Steering Committee

Pavel Brazdil	Rui Camacho
Johannes Fürnkranz	João Gama
Alípio Jorge	Joost N. Kok
Jacek Koronacki	Ramon Lopez de Mantaras
Stan Matwin	Dunja Mladenic
Tobias Scheffer	Andrzej Skowron
Myra Spiliopoulou	

Local Organization Team

Boris Cule	Guy De Pauw
Calin Garboni	Joris Gillis
Iris Hendrickx	Wim Le Page
Kim Luyckx	Michael Mampaey
Roser Morante	Adriana Prado
Koen Smets	Vincent Van Asch
Koen Van Leemput	

Area Chairs

Dimitris Achlioptas
Jean-François Boulicaut
Toon Calders
James Cussens
Tapio Elomaa
Johannes Fürnkranz
Szymon Jaroszewicz
Hillol Kargupta
Heikki Mannila
Srinivasan Parthasarathy
Lorenza Saitta
Martin Scholz
John Shawe-Taylor
Maarten van Someren
Stefan Wrobel

Roberto Bayardo
Wray Buntine
Padraig Cunningham
Ian Davidson
Martin Ester
Aris Gionis
Thorsten Joachims
Eamonn Keogh
Taneli Mielikainen
Jian Pei
Tobias Scheffer
Michèle Sébag
Antal van den Bosch
Michalis Vazirgiannis
Osmar Zaiane

Program Committee

Niall Adams
Alekh Agarwal
Charu Aggarwal
Gagan Agrawal
David Aha
Florence d'Alche-Buc
Enrique Alfonseca
Aris Anagnostopoulos
Annalisa Appice
Thierry Artieres
Yonatan Aumann
Paolo Avesani
Ricardo Baeza-Yates
Jose Balcazar
Aharon Bar Hillel
Ron Bekkerman
Bettina Berendt
Michael Berthold
Steffen Bickel
Ella Bingham
Gilles Blanchard
Hendrik Blockeel
Axel Blumenstock
Francesco Bonchi
Christian Borgelt

Karsten Borgwardt
Henrik Bostrom
Thorsten Brants
Ivan Bratko
Pavel Brazdil
Ulf Brefeld
Bjorn Bringmann
Greg Buehrer
Rui Camacho
Stephane Canu
Carlos Castillo
Deepayan Chakrabarti
Kamalika Chaudhuri
Nitesh Chawla
Sanjay Chawla
David Cheung
Alok Choudhary
Wei Chu
Alexander Clark
Chris Clifton
Ira Cohen
Antoine Cornuejols
Bruno Cremilleux
Michel Crucianu
Juan-Carlos Cubero

Gautam Das
Tijl De Bie
Luc De Raedt
Janez Demsar
Francois Denis
Giuseppe Di Fatta
Laura Dietz
Debora Donato
Dejing Dou
Kurt Driessens
Chris Drummond
Pierre Dupont
Haimonti Dutta
Pinar Duygulu-Sahin
Sašo Dzeroski
Tina Eliassi-Rad
Charles Elkan
Roberto Esposito
Floriana Esposito
Christos Faloutsos
Thomas Finley
Peter Flach
George Forman
Blaz Fortuna
Eibe Frank
Dayne Freitag
Alex Freitas
Elisa Fromont
Thomas Gaertner
Patrick Gallinari
João Gama
Dragan Gamberger
Auroop Ganguly
Gemma Garriga
Eric Gaussier
Floris Geerts
Claudio Gentile
Pierre Geurts
Amol Ghoting
Chris Giannella
Fosca Giannotti
Attilio Giordana
Mark Girolami
Derek Greene
Marko Grobelnik

Robert Grossman
Dimitrios Gunopulos
Maria Halkidi
Lawrence Hall
Jiawei Han
Tom Heskes
Alexander Hinneburg
Frank Höppner
Thomas Hofmann
Jaakko Hollmen
Tamas Horvath
Andreas Hotho
Eyke Hüllermeier
Manfred Jaeger
Sarah Jane Delany
Ruoming Jin
Alipio Jorge
Matti Kaariainen
Alexandros Kalousis
Benjamin Kao
George Karypis
Samuel Kaski
Petteri Kaski
Kristian Kersting
Ralf Klinkenberg
Adam Klivans
Jukka Kohonen
Joost Kok
Aleksander Kolcz
Stefan Kramer
Andreas Krause
Marzena Kryszkiewicz
Jussi Kujala
Pavel Laskov
Dominique Laurent
Nada Lavrac
David Leake
Chris Leckie
Philippe Leray
Carson Leung
Tao Li
Hang Li
Jinyan Li
Chih-Jen Lin
Jessica Lin

Michael Lindenbaum
 Giuseppe Lipori
 Kun Liu
 Xiaohui Liu
 Huan Liu
 Michael Madden
 Donato Malerba
 Brad Malin
 Giuseppe Manco
 Bernard Manderick
 Lluis Marquez
 Yuji Matsumoto
 Stan Matwin
 Michael May
 Prem Melville
 Rosa Meo
 Ingo Mierswa
 Pericles Mitkas
 Dunja Mladenic
 Fabian Moerchen
 Alessandro Moschitti
 Claire Nedellec
 John Nerbonne
 Jenniver Neville
 Joakim Nivre
 Richard Nock
 Alexandros Ntoulas
 Siegfried Nijssen
 Arlindo Oliveira
 Salvatore Orlando
 Miles Osborne
 Gerhard Paass
 Balaji Padmanabhan
 George Paliouras
 Themis Palpanas
 Spiros Papadimitriou
 Mykola Pechenizkiy
 Dino Pedreschi
 Ruggero Pensa
 Jose-Maria Peña
 Bernhard Pfahringer
 Petra Philips
 Enric Plaza
 Pascal Poncelet
 Doina Precup

Kai Puolamaki
 Filip Radlinski
 Shyamsundar Rajaram
 Jan Ramon
 Chotirat Ratanamahatana
 Jan Rauch
 Christophe Rigotti
 Celine Robardet
 Marko Robnik-Sikonja
 Juho Rousu
 Céline Rouveirol
 Ulrich Rueckert
 Stefan Rüping
 Hichem Sahbi
 Ansaf Salleb-Aouissi
 Taisuke Sato
 Yucel Saygin
 Bruno Scherrer
 Lars Schmidt-Thieme
 Matthias Schubert
 Marc Sebban
 Nicu Sebe
 Giovanni Semeraro
 Pierre Senellart
 Jouni Seppanen
 Benyah Shaparenko
 Arno Siebes
 Tomi Silander
 Dan Simovici
 Andrzej Skowron
 Kate Smith-Miles
 Soeren Sonnenburg
 Alessandro Sperduti
 Myra Spiliopoulou
 Ramakrishnan Srikant
 Jerzy Stefanowski
 Michael Steinbach
 Jan Struyf
 Gerd Stumme
 Tamas Sziranyi
 Domenico Talia
 Pang-Ning Tan
 Zhaohui Tang
 Evimaria Terzi
 Yannis Theodoridis

Dilys Thomas
 Kai Ming Ting
 Michalis Titsias
 Ljupco Todorovski
 Hannu Toivonen
 Marc Tommasi
 Luis Torgo
 Volker Tresp
 Panayiotis Tsaparas
 Shusaku Tsumoto
 Alexey Tsymbal
 Koen Van Hoof
 Jan Van den Bussche
 Celine Vens
 Michail Vlachos
 Gorodetsky Vladimir
 Ioannis Vlahavas
 Christel Vrain
 Jianyong Wang
 Catherine Wang
 Wei Wang
 Louis Wehenkel
 Li Wei
 Ton Weijters
 Shimon Whiteson

Jef Wijsen
 Nirmalie Wiratunga
 Anthony Wirth
 Ran Wolff
 Xintao Wu
 Michael Wurst
 Charles X. Ling
 Dong Xin
 Hui Xiong
 Dragomir Yankov
 Lexiang Ye
 Dit-Yan Yeung
 Shipeng Yu
 Chun-Nam Yu
 Yisong Yue
 Bianca Zadrozny
 Mohammed Zaki
 Gerson Zaverucha
 Changshui Zhang
 Shi Zhong
 Djamel Abdelkader Zighed
 Albrecht Zimmermann
 Jean-Daniel Zucker
 Menno van Zaanen

Additional Reviewers

Franz Acherman
 Ole Agesen
 Xavier Alvarez
 Davide Ancona
 Joaquim Aparício
 João Araújo
 Ulf Asklund
 Dharini Balasubramaniam
 Carlos Baquero
 Luís Barbosa
 Lodewijk Bergmans
 Joshua Bloch
 Noury Bouraqadi
 Johan Brichau
 Fernando Brito e Abreu
 Pim van den Broek
 Kim Bruce

Luis Caires
 Giuseppe Castagna
 Barbara Catania
 Walter Cazzola
 Shigeru Chiba
 Tal Cohen
 Aino Cornils
 Erik Corry
 Juan-Carlos Cruz
 Gianpaolo Cugola
 Pdraig Cunningham
 Christian D. Jensen
 Silvano Dal-Zilio
 Wolfgang De Meuter
 Kris De Volder
 Giorgio Delzanno
 David Detlefs

Anne Doucet
Rémi Douence
Jim Dowling
Karel Driesen
Sophia Drossopoulou
Stéphane Ducasse
Natalie Eckel
Marc Evers
Johan Fabry
Leonidas Fegaras
Luca Ferrarini
Rony Flatscher
Jacques Garrigue
Marie-Pierre Gervais
Miguel Goulão
Thomas Gschwind
Pedro Guerreiro
I. Hakki Toroslur
Görel Hedin
Christian Heide Damm
Roger Henriksson
Martin Hitz
David Holmes
James Hoover
Antony Hosking
Cengiz Icdem
Yuuji Ichisugi
Anders Ive
Hannu-Matti Järvinen
Andrew Kennedy
Graham Kirby
Svetlana Kouznetsova
Kresten Krab Thorup
Reino Kurki-Suonio
Thomas Ledoux
Yuri Leontiev
David Lorenz
Steve MacDonald
Ole Lehrmann Madsen
Eva Magnusson
Margarida Mamede
Klaus Marius Hansen
Kim Mens
Tom Mens
Isabella Merlo

Marco Mesiti
Thomas Meurisse
Mattia Monga
Sandro Morasca
M. Murat Ezbiderli
Oner N. Hamali
Hidemoto Nakada
Jacques Noye
Deniz Oguz
José Orlando Pereira
Alessandro Orso
Johan Ovlinger
Marc Pantel
Jean-François Perrot
Patrik Persson
Frédéric Peschanski
Gian Pietro Picco
Birgit Pröll
Christian Queinnec
Osmar R. Zaiane
Barry Redmond
Sigi Reich
Arend Rensink
Werner Retschitzegger
Nicolas Revault
Matthias Rieger
Mario Südholt
Paulo Sérgio Almeida
Ichiro Satoh
Tilman Schaefer
Jean-Guy Schneider
Pierre Sens
Veikko Seppänen
Magnus Steinby
Don Syme
Tarja Systä
Duane Szafron
Yusuf Tambag
Kenjiro Taura
Michael Thomsen
Sander Tichelaar
Mads Torgersen
Tom Tourwé
Arif Tumer
Ozgur Ulusoy

Werner Van Belle
Dries Van Dyck
Vasco Vasconcelos
Karsten Verelst
Cristina Videira Lopes
Juha Vihavainen

John Whaley
Mario Wolzko
Mikal Ziane
Gabi Zodik
Elena Zucca

Sponsors

We wish to express our gratitude to the sponsors of ECML PKDD 2008 for their essential contribution to the conference.



Table of Contents – Part I

Invited Talks (Abstracts)

Industrializing Data Mining, Challenges and Perspectives	1
<i>Françoise Fogelman-Soulié</i>	
From Microscopy Images to Models of Cellular Processes	2
<i>Yoav Freund</i>	
Data Clustering: 50 Years Beyond K-means	3
<i>Anil K. Jain</i>	
Learning Language from Its Perceptual Context	5
<i>Raymond J. Mooney</i>	
The Role of Hierarchies in Exploratory Data Mining	6
<i>Raghu Ramakrishnan</i>	

Machine Learning Journal Abstracts

Rollout Sampling Approximate Policy Iteration	7
<i>Christos Dimitrakakis and Michail G. Lagoudakis</i>	
New Closed-Form Bounds on the Partition Function	8
<i>Krishnamurthy Dvijotham, Soumen Chakrabarti, and Subhasis Chaudhuri</i>	
Large Margin vs. Large Volume in Transductive Learning	9
<i>Ran El-Yaniv, Dmitry Pechyony, and Vladimir Vapnik</i>	
Incremental Exemplar Learning Schemes for Classification on Embedded Devices	11
<i>Ankur Jain and Daniel Nikovski</i>	
A Collaborative Filtering Framework Based on Both Local User Similarity and Global User Similarity	12
<i>Heng Luo, Changyong Niu, Ruimin Shen, and Carsten Ullrich</i>	
A Critical Analysis of Variants of the AUC	13
<i>Stijn Vanderlooy and Eyke Hüllermeier</i>	
Improving Maximum Margin Matrix Factorization	14
<i>Markus Weimer, Alexandros Karatzoglou, and Alex Smola</i>	

Data Mining and Knowledge Discovery Journal Abstracts

Finding Reliable Subgraphs from Large Probabilistic Graphs	15
<i>Petteri Hintsanen and Hannu Toivonen</i>	
A Space Efficient Solution to the Frequent String Mining Problem for Many Databases	16
<i>Adrian Kügel and Enno Ohlebusch</i>	
The Boolean Column and Column-Row Matrix Decompositions	17
<i>Pauli Miettinen</i>	
SkyGraph: An Algorithm for Important Subgraph Discovery in Relational Graphs	18
<i>Apostolos N. Papadopoulos, Apostolos Lyritsis, and Yannis Manolopoulos</i>	
Mining Conjunctive Sequential Patterns	19
<i>Chedy Raïssi, Toon Calders, and Pascal Poncelet</i>	
Adequate Condensed Representations of Patterns	20
<i>Arnaud Soulet and Bruno Crémilleux</i>	
Two Heads Better Than One: Pattern Discovery in Time-Evolving Multi-aspect Data	22
<i>Jimeng Sun, Charalampos E. Tsourakakis, Evan Hoke, Christos Faloutsos, and Tina Eliassi-Rad</i>	

Regular Papers

TOPTMH: Topology Predictor for Transmembrane α -Helices	23
<i>Rezwan Ahmed, Huzefa Rangwala, and George Karypis</i>	
Learning to Predict One or More Ranks in Ordinal Regression Tasks . . .	39
<i>Jaime Alonso, Juan José del Coz, Jorge Díez, Oscar Luaces, and Antonio Bahamonde</i>	
Cascade RSVM in Peer-to-Peer Networks	55
<i>Hock Hee Ang, Vivekanand Gopalkrishnan, Steven C.H. Hoi, and Wee Keong Ng</i>	
An Algorithm for Transfer Learning in a Heterogeneous Environment . . .	71
<i>Andreas Argyriou, Andreas Maurer, and Massimiliano Pontil</i>	
Minimum-Size Bases of Association Rules	86
<i>José L. Balcázar</i>	
Combining Classifiers through Triplet-Based Belief Functions	102
<i>Yaxin Bi, Shengli Wu, Xuhui Shen, and Pan Xiong</i>	

An Improved Multi-task Learning Approach with Applications in Medical Diagnosis	117
<i>Jinbo Bi, Tao Xiong, Shipeng Yu, Murat Dundar, and R. Bharat Rao</i>	
Semi-supervised Laplacian Regularization of Kernel Canonical Correlation Analysis	133
<i>Matthew B. Blaschko, Christoph H. Lampert, and Arthur Gretton</i>	
Sequence Labelling SVMs Trained in One Pass	146
<i>Antoine Bordes, Nicolas Usunier, and Léon Bottou</i>	
Semi-supervised Classification from Discriminative Random Walks	162
<i>Jérôme Callut, Kevin Françoisse, Marco Saerens, and Pierre Dupont</i>	
Learning Bidirectional Similarity for Collaborative Filtering	178
<i>Bin Cao, Jian-Tao Sun, Jianmin Wu, Qiang Yang, and Zheng Chen</i>	
Bootstrapping Information Extraction from Semi-structured Web Pages	195
<i>Andrew Carlson and Charles Schafer</i>	
Online Multiagent Learning against Memory Bounded Adversaries	211
<i>Doran Chakraborty and Peter Stone</i>	
Scalable Feature Selection for Multi-class Problems	227
<i>Boris Chidlovskii and Loïc Lecerf</i>	
Learning Decision Trees for Unbalanced Data	241
<i>David A. Cieslak and Nitesh V. Chawla</i>	
Credal Model Averaging: An Extension of Bayesian Model Averaging to Imprecise Probabilities	257
<i>Giorgio Corani and Marco Zaffalon</i>	
A Fast Method for Training Linear SVM in the Primal	272
<i>Trinh-Minh-Tri Do and Thierry Artières</i>	
On the Equivalence of the SMO and MDM Algorithms for SVM Training	288
<i>Jorge López, Álvaro Barbero, and José R. Dorronsoro</i>	
Nearest Neighbour Classification with Monotonicity Constraints	301
<i>Wouter Duivesteijn and Ad Feelders</i>	
Modeling Transfer Relationships Between Learning Tasks for Improved Inductive Transfer	317
<i>Eric Eaton, Marie desJardins, and Terran Lane</i>	
Mining Edge-Weighted Call Graphs to Localise Software Bugs	333
<i>Frank Eichinger, Klemens Böhm, and Matthias Huber</i>	

Hierarchical Distance-Based Conceptual Clustering	349
<i>Ana Maria Funes, Cesar Ferri, Jose Hernandez-Orallo, and Maria Jose Ramirez-Quintana</i>	
Mining Frequent Connected Subgraphs Reducing the Number of Candidates	365
<i>Andrés Gago Alonso, José Eladio Medina Pagola, Jesús Ariel Carrasco-Ochoa, and José Fco. Martínez-Trinidad</i>	
Unsupervised Riemannian Clustering of Probability Density Functions	377
<i>Alvina Goh and René Vidal</i>	
Online Manifold Regularization: A New Learning Setting and Empirical Study	393
<i>Andrew B. Goldberg, Ming Li, and Xiaojin Zhu</i>	
A Fast Algorithm to Find Overlapping Communities in Networks	408
<i>Steve Gregory</i>	
A Case Study in Sequential Pattern Mining for IT-Operational Risk	424
<i>Valerio Grossi, Andrea Romei, and Salvatore Ruggieri</i>	
Tight Optimistic Estimates for Fast Subgroup Discovery	440
<i>Henrik Grosskreutz, Stefan Rüping, and Stefan Wrobel</i>	
Watch, Listen & Learn: Co-training on Captioned Images and Videos . . .	457
<i>Sonal Gupta, Joohyun Kim, Kristen Grauman, and Raymond Mooney</i>	
Parameter Learning in Probabilistic Databases: A Least Squares Approach	473
<i>Bernd Gutmann, Angelika Kimmig, Kristian Kersting, and Luc De Raedt</i>	
Improving k -Nearest Neighbour Classification with Distance Functions Based on Receiver Operating Characteristics	489
<i>Md. Rafiul Hassan, M. Maruf Hossain, James Bailey, and Kotagiri Ramamohanarao</i>	
One-Class Classification by Combining Density and Class Probability Estimation	505
<i>Kathryn Hempstalk, Eibe Frank, and Ian H. Witten</i>	
Efficient Frequent Connected Subgraph Mining in Graphs of Bounded Treewidth	520
<i>Tamás Horváth and Jan Ramon</i>	
Proper Model Selection with Significance Test	536
<i>Jin Huang, Charles X. Ling, Harry Zhang, and Stan Matwin</i>	

A Projection-Based Framework for Classifier Performance Evaluation . . .	548
<i>Nathalie Japkowicz, Pritika Sanghi, and Peter Tischer</i>	
Distortion-Free Nonlinear Dimensionality Reduction	564
<i>Yangqing Jia, Zheng Wang, and Changshui Zhang</i>	
Learning with $L_{q<1}$ vs L_1 -Norm Regularisation with Exponentially Many Irrelevant Features	580
<i>Ata Kabán and Robert J. Durrant</i>	
Catenary Support Vector Machines	597
<i>Kin Fai Kan and Christian R. Shelton</i>	
Exact and Approximate Inference for Annotating Graphs with Structural SVMs	611
<i>Thoralf Klein, Ulf Brefeld, and Tobias Scheffer</i>	
Extracting Semantic Networks from Text Via Relational Clustering	624
<i>Stanley Kok and Pedro Domingos</i>	
Ranking the Uniformity of Interval Pairs	640
<i>Jussi Kujala and Tapio Elomaa</i>	
Multiagent Reinforcement Learning for Urban Traffic Control Using Coordination Graphs	656
<i>Lior Kuyper, Shimon Whiteson, Bram Bakker, and Nikos Vlassis</i>	
STREAMKRIMP: Detecting Change in Data Streams	672
<i>Matthijs van Leeuwen and Arno Siebes</i>	
Author Index	689

Table of Contents – Part II

Regular Papers

Exceptional Model Mining	1
<i>Dennis Leman, Ad Feelders, and Arno Knobbe</i>	
A Joint Topic and Perspective Model for Ideological Discourse	17
<i>Wei-Hao Lin, Eric Xing, and Alexander Hauptmann</i>	
Effective Pruning Techniques for Mining Quasi-Cliques	33
<i>Guimei Liu and Limsoon Wong</i>	
Efficient Pairwise Multilabel Classification for Large-Scale Problems in the Legal Domain	50
<i>Eneldo Loza Mencía and Johannes Fürnkranz</i>	
Fitted Natural Actor-Critic: A New Algorithm for Continuous State-Action MDPs	66
<i>Francisco S. Melo and Manuel Lopes</i>	
A New Natural Policy Gradient by Stationary Distribution Metric	82
<i>Tetsuro Morimura, Eiji Uchibe, Junichiro Yoshimoto, and Kenji Doya</i>	
Towards Machine Learning of Grammars and Compilers of Programming Languages	98
<i>Keita Imada and Katsuhiko Nakamura</i>	
Improving Classification with Pairwise Constraints: A Margin-Based Approach	113
<i>Nam Nguyen and Rich Caruana</i>	
Metric Learning: A Support Vector Approach	125
<i>Nam Nguyen and Yunsong Guo</i>	
Support Vector Machines, Data Reduction, and Approximate Kernel Matrices	137
<i>XuanLong Nguyen, Ling Huang, and Anthony D. Joseph</i>	
Mixed Bregman Clustering with Approximation Guarantees	154
<i>Richard Nock, Panu Luosto, and Jyrki Kivinen</i>	
Hierarchical, Parameter-Free Community Discovery	170
<i>Spiros Papadimitriou, Jimeng Sun, Christos Faloutsos, and Philip S. Yu</i>	

A Genetic Algorithm for Text Classification Rule Induction	188
<i>Adriana Pietramala, Veronica L. Policicchio, Pasquale Rullo, and Inderbir Sidhu</i>	
Nonstationary Gaussian Process Regression Using Point Estimates of Local Smoothness	204
<i>Christian Plagemann, Kristian Kersting, and Wolfram Burgard</i>	
Kernel-Based Inductive Transfer	220
<i>Ulrich Rückert and Stefan Kramer</i>	
State-Dependent Exploration for Policy Gradient Methods	234
<i>Thomas Rückstieß, Martin Felder, and Jürgen Schmidhuber</i>	
Client-Friendly Classification over Random Hyperplane Hashes	250
<i>Shyamsundar Rajaram and Martin Scholz</i>	
Large-Scale Clustering through Functional Embedding	266
<i>Frédéric Ratle, Jason Weston, and Matthew L. Miller</i>	
Clustering Distributed Sensor Data Streams	282
<i>Pedro Pereira Rodrigues, João Gama, and Luís Lopes</i>	
A Novel Scalable and Data Efficient Feature Subset Selection Algorithm	298
<i>Sergio Rodrigues de Moraes and Alex Aussem</i>	
Robust Feature Selection Using Ensemble Feature Selection Techniques	313
<i>Yvan Saey, Thomas Abeel, and Yves Van de Peer</i>	
Effective Visualization of Information Diffusion Process over Complex Networks	326
<i>Kazumi Saito, Masahiro Kimura, and Hiroshi Motoda</i>	
Actively Transfer Domain Knowledge	342
<i>Xiaoxiao Shi, Wei Fan, and Jiangtao Ren</i>	
A Unified View of Matrix Factorization Models	358
<i>Ajit P. Singh and Geoffrey J. Gordon</i>	
Parallel Spectral Clustering	374
<i>Yangqiu Song, Wen-Yen Chen, Hongjie Bai, Chih-Jen Lin, and Edward Y. Chang</i>	
Classification of Multi-labeled Data: A Generative Approach	390
<i>Andreas P. Streich and Joachim M. Buhmann</i>	
Pool-Based Agnostic Experiment Design in Linear Regression	406
<i>Masashi Sugiyama and Shinichi Nakajima</i>	
Distribution-Free Learning of Bayesian Network Structure	423
<i>Xiaohai Sun</i>	

Assessing Nonlinear Granger Causality from Multivariate Time Series	440
<i>Xiaohai Sun</i>	
Clustering Via Local Regression	456
<i>Jun Sun, Zhiyong Shen, Hui Li, and Yidong Shen</i>	
Decomposable Families of Itemsets	472
<i>Nikolaj Tatti and Hannes Heikinheimo</i>	
Transferring Instances for Model-Based Reinforcement Learning	488
<i>Matthew E. Taylor, Nicholas K. Jong, and Peter Stone</i>	
A Simple Model for Sequences of Relational State Descriptions	506
<i>Ingo Thon, Niels Landwehr, and Luc De Raedt</i>	
Semi-supervised Boosting for Multi-Class Classification	522
<i>Hamed Valizadegan, Rong Jin, and Anil K. Jain</i>	
A Joint Segmenting and Labeling Approach for Chinese Lexical Analysis	538
<i>Xinhao Wang, Jiazhong Nie, Dingsheng Luo, and Xihong Wu</i>	
Transferred Dimensionality Reduction	550
<i>Zheng Wang, Yangqiu Song, and Changshui Zhang</i>	
Multiple Manifolds Learning Framework Based on Hierarchical Mixture Density Model	566
<i>Xiaoxia Wang, Peter Tiño, and Mark A. Fardal</i>	
Estimating Sales Opportunity Using Similarity-Based Methods	582
<i>Sholom M. Weiss and Nitin Indurkha</i>	
Learning MDP Action Models Via Discrete Mixture Trees	597
<i>Michael Wynkoop and Thomas Dietterich</i>	
Continuous Time Bayesian Networks for Host Level Network Intrusion Detection	613
<i>Jing Xu and Christian R. Shelton</i>	
Data Streaming with Affinity Propagation	628
<i>Xiangliang Zhang, Cyril Furtlehner, and Michèle Sebag</i>	
Semi-supervised Discriminant Analysis Via CCCP	644
<i>Yu Zhang and Dit-Yan Yeung</i>	

Demo Papers

A Visualization-Based Exploratory Technique for Classifier Comparison with Respect to Multiple Metrics and Multiple Domains	660
<i>Rocío Alaiz-Rodríguez, Nathalie Japkowicz, and Peter Tischer</i>	

<i>Pleiades: Subspace Clustering and Evaluation</i>	666
<i>Ira Assent, Emmanuel Müller, Ralph Krieger, Timm Jansen, and Thomas Seidl</i>	
<i>SEDiL: Software for Edit Distance Learning</i>	672
<i>Laurent Boyer, Yann Esposito, Amaury Habrard, Jose Oncina, and Marc Sebban</i>	
<i>Monitoring Patterns through an Integrated Management and Mining Tool</i>	678
<i>Evangelos E. Kotsifakos, Irene Ntoutsi, Yannis Vrahoritis, and Yannis Theodoridis</i>	
<i>A Knowledge-Based Digital Dashboard for Higher Learning Institutions</i>	684
<i>Wan Maseri Binti Wan Mohd, Abdullah Embong, and Jasni Mohd Zain</i>	
<i>SINDBAD and SiQL: An Inductive Database and Query Language in the Relational Model</i>	690
<i>Jörg Wicker, Lothar Richter, Kristina Kessler, and Stefan Kramer</i>	
Author Index	695