

Lecture Notes in Artificial Intelligence 4213

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Johannes Fürnkranz Tobias Scheffer
Myra Spiliopoulou (Eds.)

Knowledge Discovery in Databases: PKDD 2006

10th European Conference on Principle and Practice
of Knowledge Discovery in Databases
Berlin, Germany, September 18-22, 2006
Proceedings

Series Editors

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

Volume Editors

Johannes Fürnkranz
Technische Universität Darmstadt
Fachbereich Informatik
Hochschulstraße 10, 64289 Darmstadt, Germany
E-mail: juffi@ke.informatik.tu-darmstadt.de

Tobias Scheffer
Humboldt-Universität zu Berlin
Institut für Informatik
Unter den Linden 6, 10099 Berlin, Germany
E-mail: scheffer@informatik.hu-berlin.de

Myra Spiliopoulou
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Universitätsplatz 2, 39016 Magdeburg, Germany
E-mail: myra@iti.cs.uni-magdeburg.de

Library of Congress Control Number: 2006932511

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

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-45374-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-45374-1 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 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11871637 06/3142 5 4 3 2 1 0

Preface

The two premier annual European conferences in the areas of machine learning and data mining have been collocated ever since the joint conference in Freiburg, Germany, 2001. The European Conference on Machine Learning was established 20 years ago, when the first European Working Session on Learning was held in Orsay, France, in 1986. The conference is growing, and is more lively than ever. The European Conference on Principles and Practice of Knowledge Discovery in Databases celebrates its tenth anniversary; the first PKDD took place in 1997 in Trondheim, Norway. Over the years, the ECML/PKDD series has evolved into one of the largest and most selective international conferences in these areas, the only one that provides a common forum for the two closely related fields. In 2006, the 6th collocated ECML/PKDD took place during September 18-22, when the Humboldt-Universität zu Berlin hosted the 17th European Conference on Machine Learning (ECML) and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD).

The successful model of a hierarchical reviewing process that was introduced last year for the ECML/PKDD 2005 in Porto has been taken over in 2006. We nominated 32 Area Chairs, each of them responsible for several closely related research topics. Suitable areas were selected on the basis of the submission statistics for ECML/PKDD 2005 to ensure a proper load balance among the Area Chairs. For the first time, a joint Program Committee was nominated for the two conferences, consisting of 280 renowned researchers, mostly proposed by the Area Chairs. This joint PC, the largest of the series to date, allowed us to exploit synergies and deal competently with topic overlaps between ECML and PKDD.

ECML/PKDD 2006 received 564 full paper submissions that entered the reviewing process. The submissions were manually assigned to the Area Chairs, who coordinated the reviewers thereafter. Reviewer assignment was based on bidding with CyberChairPRO, as in the previous years. With very few exceptions, every submission was reviewed by three PC members. Based on these reviews, on feedback from the authors, and on discussions among the reviewers, the Area Chairs provided a recommendation for each paper. Continuing the tradition of previous events in the series, we accepted full papers for oral presentation and short papers for poster presentation. The final decision was made by us based on the recommendations of the Area Chairs. We selected 46 full papers and 36 short papers for ECML, and 36 full papers and 26 short papers for PKDD. The acceptance rate for full papers is 14.5% and the overall acceptance rate is 25.5%, in accordance with the high-quality standards of the conference series. Next to the paper and poster sessions, ECML/PKDD 2006 also featured five invited talks, ten workshops, seven tutorials and the ECML/PKDD discovery challenge.

We distinguished eight outstanding contributions; the awards were generously sponsored by the *Machine Learning Journal* and the *KD-Ubiq network*.

ECML Best Paper: Quoc Le, Alex Smola, Thomas Gärtner, Yasemin Altun: *Transductive Gaussian Process Regression with Automatic Model Selection.*

PKDD Best Paper: Pauli Miettinen, Taneli MielikÄäen, Aristides Gionis, Gautam Das, Heikki Mannila: *The Discrete Basis Problem.*

ECML Best Student Paper: Bernd Gutmann and Kristian Kersting: *TildeCRF. Conditional Random Fields for Logical Sequences.*

PKDD Best Student Paper: Arik Friedmann, Assaf Schuster, Ran Wolff: *k-Anonymous Decision Tree Induction.*

ECML Innovative Contribution: Alexander Clark, Christophe Costa Florencio, Chris Watkins: *Languages as Hyperplanes: Grammatical Inference with String Kernels.*

PKDD Innovative Application: Herna Viktor, Eric Paquet, Hongyu Guo: *Measuring to Fit: Virtual Tailoring Through Cluster Analysis and Classification.*

The *ECML/PKDD Best Presentation* and the *ECML/PKDD Best Poster Presentation* awards were elected by participants of the conference.

This year's *Discovery Challenge* focused on personalized spam filtering and generalization across related learning tasks. Steffen Bickel organized the challenge; 26 teams participated. For task A, three teams achieved a first rank: Khurram Nazir Junejo, Mirza Muhammad Yousaf, and Asim Karim; Bernhard Pfahringer; and Kushagra Gupta, Vikrant Chaudhary, Nikhil Marwah, and Chirag Taneja. Task B was won by Gordon Cormack. The solution of Bernhard Pfahringer was distinguished with the Creativity Award.

We are indebted to the Area Chairs, Program Committee members and external reviewers for their effort and engagement in making a rich but selective scientific program for ECML/PKDD. Special thanks go to those reviewers that helped with additional reviews at very short notice to assist with at difficult decisions. We further thank our two workshop and tutorial chairs Tapio Elomaa and Bart Goethals for selecting and coordinating the ten workshop and seven tutorial events that accompanied the conference; the workshop organizers, tutorial presenters, and the organizers of the discovery challenge; Richard van de Stadt and CyberChairPRO for competent and flexible support; the Local Organizing Committee; and all other people that contributed to the organization of this event. Finally, we are grateful to the the Steering Committee and the ECML/PKDD community that entrusted us with the organization of ECML/PKDD 2006.

Most of all, however, we would like to thank all the authors who honored us by submitting their work to this conference, thereby facilitating the success of this event.

Sponsors

We wish to express our gratitude to our sponsors for their great contributions to the conference. We wish to thank Google for featuring the Google ECML Poster Reception and providing ten Student Travel Awards; the Humboldt-Universität zu Berlin for providing the conference venue; the German Science Foundation DFG for supporting all invited speakers; KD-Ubiq for supporting the PKDD Poster Reception and European Projects Poster Reception, four Student Travel Awards, and the Best Paper Awards; the European Office of Aerospace Research and Development (EOARD), Air Force Office of Scientific Research, United States Air Force Research Laboratory for generous financial support; Strato AG for providing the awards to the winners of the Discovery Challenge; the Pascal Network of Excellence and IBM for financial support; the *Machine Learning Journal* for supporting the Student Best Paper Awards.

Google™



Deutsche
Forschungsgemeinschaft

DFG



*Machine
Learning*

AFOSR/EOARD support is not intended to express or imply endorsement by the U.S. Federal Government.

Organization

Program Chairs

Johannes Fürnkranz (Technische Universität Darmstadt)
Tobias Scheffer (Humboldt-Universität zu Berlin)
Myra Spiliopoulou (Otto-von-Guericke-Universität Magdeburg)

Local Chairs

Andreas Nürnberger (Otto-von-Guericke-Universität Magdeburg)
Tobias Scheffer (Humboldt-Universität zu Berlin)

Workshop and Tutorial Chairs

Tapio Elomaa (Tampere University of Technology)
Bart Goethals (University of Antwerp)

Discovery Challenge Chair

Steffen Bickel (Humboldt-Universität zu Berlin)

Poster and Fashion Designer, Webmaster

Michael Brückner (Humboldt-Universität zu Berlin)

Local Arrangements Team

Steffen Bickel	Ulf Brefeld
Michael Brückner	Isabel Drost
Thoralf Klein	

Steering Committee

Hendrik Blockeel	Jean-François Boulicaut
Pavel Brazdil	Rui Camacho
Florian Esposito	João Gama
Dragan Gamberger	Fosca Gianotti
Alípio Jorge	Nada Lavrač
Dino Pedreschi	Ljupčo Todorovski
Luís Torgo	

Area Chairs

Hendrik Blockeel
Soumen Chakrabarti
Pádraig Cunningham
Kurt Driessens
Peter Flach
Thomas Gärtner
Bart Goethals
George Karypis
Stefan Kramer
Katharina Morik
Bernhard Pfahringer
Carl Rasmussen
Dale Schuurmans
Csaba Szepesvári
Luís Torgo
Vishy Vishwanathan

Henrik Boström
Olivier Chapelle
James Cussens
Ronen Feldman
Eibe Frank
João Gama
Thomas Hofmann
Kristian Kersting
Stan Matwin
Klaus Obermayer
Luc De Raedt
Lorenza Saitta
Jaideep Srivastava
Hannu Toivonen
Volker Tresp
Mohammed Zaki

Program Committee

James Abello
Douglas Aberdeen
Gediminas Adomavicius
Charu Aggarwal
Eugene Agichtein
Jesus S. Aguilar-Ruiz
Erick Alphonse
Yasemin Altun
Massih Amini
Aijun An
Josep-Lluís Arcos
Yonatan Aumann
Paolo Avesani
Antonio Bahamonde
David Barber
Roberto Bayardo
Bettina Berendt
Michael R. Berthold
Elisa Bertino
Fernando Berzal
Steffen Bickel
Francesco Bonchi
Gianluca Bontempi
Remco Bouckaert

Jean-François Boulicaut
Thorsten Brants
Mikio Braun
Pavel Brazdil
Ulf Brefeld
Derek Bridge
Björn Bringmann
Klaus Brinker
Wray Buntine
Christian Böhm
Tiberio Caetano
Toon Calders
Rui Camacho
Barbara Catania
Nicolò Cesa-Bianchi
Deepayan Chakrabarti
Sharma Chakravarthy
LiWu Chang
David Cheung
Amanda Clare
Ian Cloete
Antoine Cornuéjols
Koby Crammer
Susan Craw

Daniel Cremers
Bruno Cremilleux
Walter Daelemans
Ernst Damien
Hal Daume III
Ian Davidson
Luc Dehaspe
Olivier Delalleau
Janez Demšar
François Denis
Chris Ding
Chabane Djeraba
Sašo Džeroski
Tina Eliassi-Rad
Tapio Elomaa
Floriana Esposito
Roberto Esposito
Martin Ester
Wei Fan
Ad Feelders
Arthur Flexer
George Forman
Matthias Franz
Paolo Frasconi
Glenn Fung
Linda C. van der Gaag
Dragan Gamberger
Jean-Gabriel Ganascia
Minos Garofalakis
Floris Geerts
Pierre Geurts
Ali Ghodsi
Joydeep Ghosh
Fosca Giannotti
Aristides Gionis
Attilio Giordana
Christophe Giraud-Carrier
Mark Girolami
David Gondek
Gunter Grieser
Marko Grobelnik
Valerie Guralnik
Mark Hall
Howard Hamilton
Eui-Hong Han

Markus Hegland
Jose Hernandez-Orallo
Gerhard Heyer
Colin de la Higuera
Melanie Hilario
Alexander Hinneburg
Susanne Hoche
Achim Hoffmann
Jaakko Hollmén
John H. Holmes
Geoffrey Holmes
Tamas Horvath
Andreas Hotho
San-Yih Hwang
Frank Höppner
Eyke Hüllermeier
Nitin Indurkha
Iñaki Inza
Manfred Jaeger
Szymon Jaroszewicz
Edwin de Jong
Alípio Jorge
Alexandros Kalousis
Jaz Kandola
Hillol Kargupta
Andreas Karwath
Samuel Kaski
Motoaki Kawanabe
Eamonn Keogh
Latifur Khan
Ross King
Mika Klemettinen
Ralf Klinkenberg
Arno Knobbe
Jens Kohlmorgen
Igor Kononenko
Adam Kowalczyk
Miroslav Kubat
Nicolas Lachiche
Michail Lagoudakis
Pedro Larrañaga
Pavel Laskov
Mark Last
Dominique Laurent
Nada Lavrač

Ulf Leser
Jure Leskovec
Christina Leslie
Ee-Peng Lim
Bing Liu
John Lloyd
Huma Lodhi
Jose A. Lozano
Ulrike von Luxburg
Donato Malerba
Suresh Manandhar
Giuseppe Manco
Heikki Mannila
Yannis Manolopoulos
Dragos Margineantu
Yuji Matsumoto
Michael May
Vasileios Megalooikonomou
Wagner Meira Jr.
Prem Melville
Rosa Meo
Taneli Mielikäinen
Rada Mihalcea
Brian Milch
Dunja Mladenic
Bamshad Mobasher
Klaus-Robert Müller
Hector Muñoz-Avila
Alexandros Nanopoulos
Olfa Nasraoui
Claire Nedellec
Kee Siong Ng
Siegfried Nijssen
Mahesan Niranjan
Richard Nock
Ann Nowe
Andreas Nürnberger
Arlindo Oliveira
Manfred Opper
Carlos Ordonez
Miles Osborne
Martijn van Otterlo
Gerhard Paaß
Balaji Padmanabhan
Georgios Paliouras

Spiros Papadimitriou
Srinivasan Parthasarathy
Andrea Passerini
Dino Pedreschi
Jian Pei
Lourdes Peña
Jean-Marc Petit
Johann Petrak
Enric Plaza
William M. Pottenger
Doina Precup
Stephen Pulman
Joaquin Quiñonero-Candela
Naren Ramakrishnan
Jan Ramon
Rajeev Rastogi
Francesco Ricci
Christophe Rigotti
Gilbert Ritschard
Céline Robardet
Marko Robnik-Sikonja
Romer Rosales
Volker Roth
Juho Rousu
Céline Rouveirol
Stefan Rüping
Ulrich Rückert
Cordelia Schmid
Martin Scholz
Nic Schraudolph
Anton Schwaighofer
Michele Sebag
Marc Sebban
Matthias Seeger
Bernhard Seeger
Giovanni Semeraro
Petteri Sevón
Ayumi Shinohara
Arno Siebes
Dan Simovici
Neil Smalheiser
Padhraic Smyth
Carlos Soares
Stephen Soderland
Maarten van Someren

Alessandro Sperduti
 Eduardo Spinosa
 Nicolas Spyratos
 Michael Steinbach
 Jan Struyf
 Gerd Stumme
 Einoshin Suzuki
 Vojtech Svatek
 Marcin Szczuka
 Prasad Tadepalli
 Pang-Ning Tan
 Tao Tao
 Takao Terano
 Henry Tirri
 Ljupco Todorovski
 Ioannis Tsochantaridis
 Karl Tuyls
 Michalis Vazirgiannis
 Katja Verbeeck
 Jean-Philippe Vert
 Rene Vidal
 Jianyong Wang
 Wei Wang
 Ke Wang
 Jason Wang
 Haixun Wang

Takashi Washio
 Louis Wehenkel
 Sholom Weiss
 Gerhard Widmer
 Marco Wiering
 Eric Wiewiora
 Ole Winther
 Nirmalie Wiratunga
 Stefan Wrobel
 Hui Xiong
 Jiong Yang
 Ying Yang
 Philip Yu
 Kai Yu
 Bianca Zadrozny
 Gerson Zaverucha
 Filip Zelezny
 Thomas Zeugmann
 ChengXiang Zhai
 Byoung-Tak Zhang
 Ying Zhao
 Dengyong Zhou
 Xiaojin Zhu
 Alexander Zien
 Jean-Daniel Zucker
 Blaž Zupan

Additional Reviewers

Sreangsu Acharyya
 Elke Aichert
 Raman Adaikkalavan
 Fabio Aioli
 Florence d'Alché-Buc
 Claudia d'Amato
 Bill Andreopoulos
 Annalisa Appice
 Eva Armengol
 Stella Asimwe
 Anneleen Van Assche
 Maurizio Atzori
 Korinna Bade
 Gökhan H. Bakır
 Pierpaolo Basile

Matthew Beal
 Gurkan Bebek
 Aditya Belapurkar
 Margherita Berardi
 Jürgen Beringer
 Marc Bernard
 Kanishka Bhaduri
 Marenglen Biba
 Hanczar Blaise
 Gilles Blanchard
 Yann-aël Le Borgne
 Karsten Borgwardt
 Marco Botta
 Dominique Bouthinon
 Janez Brank

Michael Brückner
Robert D. Burbidge
Emma Byrne
Nicolas Cebon
Michelangelo Ceci
Eugenio Cesario
Hasan E. Cetingul
Sutanu Chakraborti
Hari Chelladurai
Min Chen
Haibin Cheng
Hong Cheng
Li Cheng
Yun Chi
Silvia Chiappa
Thiago Cordeiro
Gianni Costa
Marco Cristo
Tom Croonenborghs
Vassilis Cutsuridis
Wisam Dakka
Souptik Datta
Sandra De Amo
Fabien De Marchi
Marco Degemmis
Evangelos Dellis
Nicola Di Mauro
Isabel Drost
Norberto Díaz Díaz
Lauri Eronen
Vicent Estruch
Nicola Fanizzi
Francisco J. Ferrer-Troyano
Stefano Ferilli
Cèsar Ferri
Frédéric Flouvat
Francesco Folino
Nuno Fonseca
Blaz Fortuna
Murat Ganiz
Jing Gao
Dimitrios Gavrilis
Peter Vincent Gehler
Liqiang Geng
Pierre Gerard

Atiyeh Ghoreyshi
Amol Ghoting
Arnaud Giacometti
Chris Giannella
Aris Gkoulalas-Divanis
Robby Goetschalckx
Alvina Goh
Ricardo Grau
Steve Greenwald
Gunjan Gupta
Simon Günter
Benjamin Haibe-Kains
Maria Halkidi
Hongxing He
Matthias Hein
Christoph Heinz
Corneliu Henegar
Katherine G. Herbert
Kirsten Hildrum
Petteri Hintsanen
Wai-Shing Ho
Meng Hu
Ming Hua
Jiayuan Huang
Yi Huang
Alex Ihler
Jaikrishna
Christopher D. Janneck
Robert Jäschke
Tao-Yuan Jen
Antonio Jimeno
Frederic Jurie
Dmitry Kamenetsky
Murat Kantarcioglu
Ioannis Karydis
Balakumar Kendai
Wolf Kienzle
David H. Kim
Seyoung Kim
Jee-Hyub Kim
Kwang In Kim
Jiří Kléma
Ioannis Klafaptis
Christine Körner
Ville Kononen

Stasinou Konstantopoulos	Aritz Pérez
Despina Kontos	Ignazio Palmisano
Arne Koopman	Elias Pampalk
Petra Kralj	Maarten Peeters
Chase Krumpelman	Jaakko Peltonen
Jussi Kujala	Michael Perrow
Matjaž Kukar	Sergios Petridis
Peer Kröger	Viet Phan-Luong
Peter Kunath	Tadek Pietraszek
Niels Landwehr	Aloisio Carlos de Pina
Maggie Man Ki Lau	Conrad Plake
Anne Laurent	Claudia Plant
Quoc V Le	Mannes Poel
Steven Lemm	Subhesh Pradhan
Shenzhi Li	Pavel Praks
Ting Li	Kai Puolamäki
Oriana Licchelli	Gunnar Rättsch
Marina Litvak	M. Jose Ramirez-Quintana
Alexander Liu	S. S. Ravi
Kun Liu	Avinash Ravichandran
Yang Liu	Simon Rawles
George Loizou	Matthias Renz
Pasquale Lops	Chiara Renso
Alicia Troncoso Lora	Konrad Rieck
Marc Q. Ma	Stefan Riezler
Patrick Marcel	Carsten Riggelsen
Keith Marsolo	Salvatore Rinzivillo
Eric Martin	Pedro Rodrigues
Elio Masciari	Domingo S. Rodríguez-Baena
Stewart Massie	Thierson Rosa
Giuseppe M. Mazzeo	Michel de Rougemont
Oscar Meruvia Pastor	Vishesh Sahu
Kapila Moonesinghe	Luka Šajn
Kawanabe Motoaki	Sagar Savla
Fernando Mourão	Leander Schietgat
Rahman Mukras	Christoph Schmitz
Mirco Nanni	Timon Schroeter
Jan Nemrava	Matthias Schubert
Isabel A. Nepomuceno-Chamorro	Jerry Scripps
David Newman	Tomi Silander
Radu Stefan Niculescu	Milan Simunek
Blaž Novak	Dheeraj Singaraju
Cheng Soon Ong	Ksenia Shalnova
Amandine Orecchioni	Kazumi Slott
Matthew Eric Otey	Diego Sona

Bin Song
Le Song
Xiaodan Song
Soeren Sonnenburg
Arnaud Soulet
Junilda Spirollari
Harald Steck
Sebastian Stober
Petr Strossa
Jimeng Sun
Peter Sunehag
Aditya Telang
Choon Hui Teo
Ivan Titov
Igor Trajkovski
George Tsatsaronis
Mihalis Tsoukalos
Miroslav Vacura
Hamed Valizadegan
Antonio Varlaro
Sriharsha Veeramachaneni
Arvind Venkatachalam
Jarkko Venna
Jakob Verbeek
Vassilis Verykios
Akrivi Vlachou

Peter Vrancx
Christian James Walder
Qian Wan
Chao Wang
Lei Wang
Qiang Wang
Xuanhui Wang
Li Wei
Joost van de Weijer
Dietrich Wetschereck
Adam Woznica
Junjie Wu
Mingrui Wu
Xiaopeng Xi
Keisuke Yamazaki
Hui Yang
Xiaoning Yang
Dragomir Yankov
Jin Yu
Shipeng Yu
Monika Zakova
Shijie Zhang
Xinhua Zhang
Bin Zhou
Xingquan Zhu
Arthur Zimek

Table of Contents

Invited Talks

On Temporal Evolution in Data Streams	1
<i>Charu C. Aggarwal</i>	
The Future of CiteSeer: CiteSeer ^x	2
<i>C. Lee Giles</i>	
Learning to Have Fun	3
<i>Jonathan Schaeffer</i>	
Winning the DARPA Grand Challenge	4
<i>Sebastian Thrun</i>	
Challenges of Urban Sensing	5
<i>Henry Tirri</i>	

Long Papers

SD-Map – A Fast Algorithm for Exhaustive Subgroup Discovery	6
<i>Martin Atzmueller, Frank Puppe</i>	
Decision Trees for Hierarchical Multilabel Classification: A Case Study in Functional Genomics	18
<i>Hendrik Blockeel, Leander Schietgat, Jan Struyf, Sašo Džeroski, Amanda Clare</i>	
Clustering Scientific Literature Using Sparse Citation Graph Analysis	30
<i>Levent Boilelli, Seyda Ertekin, C. Lee Giles</i>	
VOGUE: A Novel Variable Order-Gap State Machine for Modeling Sequences	42
<i>Bouchra Bouqata, Christopher D. Carothers, Boleslaw K. Szymanski, Mohammed J. Zaki</i>	
Don't Be Afraid of Simpler Patterns	55
<i>Björn Bringmann, Albrecht Zimmermann, Luc De Raedt, Siegfried Nijssen</i>	

An Adaptive Prequential Learning Framework for Bayesian Network Classifiers	67
<i>Gladys Castillo, João Gama</i>	
Adaptive Active Classification of Cell Assay Images	79
<i>Nicolas Cebron, Michael R. Berthold</i>	
Learning Parameters in Entity Relationship Graphs from Ranking Preferences	91
<i>Soumen Chakrabarti, Alekh Agarwal</i>	
Detecting Fraudulent Personalities in Networks of Online Auctioneers	103
<i>Duen Hornq Chau, Shashank Pandit, Christos Faloutsos</i>	
Measuring Constraint-Set Utility for Partitional Clustering Algorithms	115
<i>Ian Davidson, Kiri L. Wagstaff, Sugato Basu</i>	
Discovery of Interesting Regions in Spatial Data Sets Using Supervised Clustering	127
<i>Christoph F. Eick, Banafsheh Vaezian, Dan Jiang, Jing Wang</i>	
Optimal String Mining Under Frequency Constraints	139
<i>Johannes Fischer, Volker Heun, Stefan Kramer</i>	
k -Anonymous Decision Tree Induction	151
<i>Arik Friedman, Assaf Schuster, Ran Wolff</i>	
Closed Sets for Labeled Data	163
<i>Gemma C. Garriga, Petra Kralj, Nada Lavrač</i>	
Finding Trees from Unordered 0–1 Data	175
<i>Hannes Heikinheimo, Heikki Mannila, Jouni K. Seppänen</i>	
Web Communities Identification from Random Walks	187
<i>Jiayuan Huang, Tingshao Zhu, Dale Schuurmans</i>	
Information Marginalization on Subgraphs	199
<i>Jiayuan Huang, Tingshao Zhu, Russell Greiner, Dengyong Zhou, Dale Schuurmans</i>	
Why Does Subsequence Time-Series Clustering Produce Sine Waves?	211
<i>Tsuyoshi Idé</i>	

Transductive Learning for Text Classification Using Explicit Knowledge Models	223
<i>Georgiana Ifrim, Gerhard Weikum</i>	
Exploring Multiple Communities with Kernel-Based Link Analysis	235
<i>Takahiko Ito, Masashi Shimbo, Daichi Mochihashi, Yuji Matsumoto</i>	
Distribution Rules with Numeric Attributes of Interest	247
<i>Alípio M. Jorge, Paulo J. Azevedo, Fernando Pereira</i>	
Tractable Models for Information Diffusion in Social Networks	259
<i>Masahiro Kimura, Kazumi Saito</i>	
Efficient Spatial Classification Using Decoupled Conditional Random Fields	272
<i>Chi-Hoon Lee, Russell Greiner, Osmar Zaiane</i>	
Group SAX: Extending the Notion of Contrast Sets to Time Series and Multimedia Data	284
<i>Jessica Lin, Eamonn Keogh</i>	
An Attacker's View of Distance Preserving Maps for Privacy Preserving Data Mining	297
<i>Kun Liu, Chris Giannella, Hillol Kargupta</i>	
A Scalable Distributed Stream Mining System for Highway Traffic Data	309
<i>Ying Liu, Alok Choudhary, Jianhong Zhou, Ashfaq Khokhar</i>	
K-Landmarks: Distributed Dimensionality Reduction for Clustering Quality Maintenance	322
<i>Panagis Magdalinos, Christos Doulkeridis, Michalis Vazirgiannis</i>	
The Discrete Basis Problem	335
<i>Pauli Miettinen, Taneli Mielikäinen, Aristides Gionis, Gautam Das, Heikki Mannila</i>	
Evaluation of Summarization Schemes for Learning in Streams	347
<i>Alec Pawling, Nitesh V. Chawla, Amitabh Chaudhary</i>	
Efficient Mining of Correlation Patterns in Spatial Point Data	359
<i>Marko Salmenkivi</i>	

Improving Functional Modularity in Protein-Protein Interactions
 Graphs Using Hub-Induced Subgraphs 371
*Duygu Ucar, Sitaram Asur, Umit Catalyurek,
 Srinivasan Parthasarathy*

Refining Aggregate Conditions in Relational Learning 383
Celine Vens, Jan Ramon, Hendrik Blockeel

Measuring to Fit: Virtual Tailoring Through Cluster Analysis
 and Classification 395
Herna L. Viktor, Eric Paquet, Hongyu Guo

RIVA: Indexing and Visualization of High-Dimensional Data Via
 Dimension Reorderings 407
*Michail Vlachos, Spiros Papadimitriou, Zografoula Vagena,
 Philip S. Yu*

Distributed Subgroup Mining 421
Michael Wurst, Martin Scholz

Network Flow for Collaborative Ranking 434
Ziming Zhuang, Silviu Cucerzan, C. Lee Giles

Short Papers

Finding Hierarchies of Subspace Clusters 446
*Elke Aichtert, Christian Böhm, Hans-Peter Kriegel, Peer Kröger,
 Ina Müller-Gorman, Arthur Zimek*

Integrating Pattern Mining in Relational Databases 454
Toon Calders, Bart Goethals, Adriana Prado

Discovering Patterns in Real-Valued Time Series 462
Joe Catalano, Tom Armstrong, Tim Oates

Classification of Dementia Types from Cognitive Profiles Data 470
*Giorgio Corani, Chris Edgar, Isabelle Marshall, Keith Wesnes,
 Marco Zaffalon*

When Efficient Model Averaging Out-Performs Boosting and Bagging . . . 478
Ian Davidson, Wei Fan

Peak-Jumping Frequent Itemset Mining Algorithms 487
Nele Dexters, Paul W. Purdom, Dirk Van Gucht

Autonomous Visualization	495
<i>Khalid El-Arini, Andrew W. Moore, Ting Liu</i>	
Naive Bayes for Text Classification with Unbalanced Classes.....	503
<i>Eibe Frank, Remco R. Bouckaert</i>	
Knowledge-Conscious Exploratory Data Clustering	511
<i>Amol Ghoting, Srinivasan Parthasarathy</i>	
On the Lower Bound of Reconstruction Error for Spectral Filtering Based Privacy Preserving Data Mining	520
<i>Songtao Guo, Xintao Wu, Yingjiu Li</i>	
Frequent Pattern Discovery Without Binarization: Mining Attribute Profiles	528
<i>Attila Gyenesei, Ralph Schlapbach, Etzard Stolte, Ulrich Wagner</i>	
Efficient Name Disambiguation for Large-Scale Databases	536
<i>Jian Huang, Seyda Ertekin, C. Lee Giles</i>	
Adaptive Segmentation-Based Symbolic Representations of Time Series for Better Modeling and Lower Bounding Distance Measures	545
<i>Bernard Huguency</i>	
A Feature Generation Algorithm for Sequences with Application to Splice-Site Prediction	553
<i>Rezarta Islamaj, Lise Getoor, W. John Wilbur</i>	
Discovering Image-Text Associations for Cross-Media Web Information Fusion	561
<i>Tao Jiang, Ah-Hwee Tan</i>	
Mining Sequences of Temporal Intervals	569
<i>Steffen Kempe, Jochen Hipp</i>	
Pattern Teams.....	577
<i>Arno J. Knobbe, Eric K.Y. Ho</i>	
Compression Picks Item Sets That Matter	585
<i>Matthijs van Leeuwen, Jilles Vreeken, Arno Siebes</i>	
Discovering Overlapping Communities of Named Entities	593
<i>Xin Li, Bing Liu, Philip S. Yu</i>	
Closed Non-derivable Itemsets	601
<i>Juho Muhonen, Hannu Toivonen</i>	

Learning a Distance Metric for Object Identification Without
Human Supervision 609
Satoshi Oyama, Katsumi Tanaka

Towards Association Rules with Hidden Variables 617
Ricardo Silva, Richard Scheines

A Data Mining Approach to the Joint Evaluation of Field
and Manufacturing Data in Automotive Industry 625
Christian Manuel Strobel, Tomas Hrycej

Incremental Aspect Models for Mining Document Streams 633
Arun C. Surendran, Swrit Sra

Learning Approximate MRFs from Large Transaction Data 641
Chao Wang, Srinivasan Parthasarathy

Similarity Search for Multi-dimensional NMR-Spectra
of Natural Products 650
Karina Wolfram, Andrea Porzel, Alexander Hinneburg

Author Index 659