

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Sašo Džeroski Jan Struyf (Eds.)

Knowledge Discovery in Inductive Databases

5th International Workshop, KDID 2006

Berlin, Germany, September 18, 2006

Revised Selected and Invited Papers

Volume Editors

Sašo Džeroski
Jožef Stefan Institute
Department of Knowledge Technologies
Jamova 39, 1000 Ljubljana, Slovenia
E-mail: saso.dzeroski@ijs.si

Jan Struyf
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A, 3001 Leuven, Belgium
E-mail: jan.struyf@cs.kuleuven.be

Library of Congress Control Number: 2007937944

CR Subject Classification (1998): H.2, I.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-540-75548-9 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-75548-7 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 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12171675 06/3180 5 4 3 2 1 0

Preface

The 5th International Workshop on Knowledge Discovery in Inductive Databases (KDID 2006) was held on September 18, 2006 in Berlin, Germany, in conjunction with ECML/PKDD 2006: The 17th European Conference on Machine Learning (ECML) and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD).

Inductive databases (IDBs) represent a database view on data mining and knowledge discovery. IDBs contain not only data, but also generalizations (patterns and models) valid in the data. In an IDB, ordinary queries can be used to access and manipulate data, while inductive queries can be used to generate (mine), manipulate, and apply patterns. In the IDB framework, patterns become “first-class citizens”, and KDD becomes an extended querying process in which both the data and the patterns/models that hold in the data are queried.

The IDB framework is appealing as a general framework for data mining, because it employs declarative queries instead of ad-hoc procedural constructs. As declarative queries are often formulated using constraints, inductive querying is closely related to constraint-based data mining. The IDB framework is also appealing for data mining applications, as it supports the entire KDD process, i.e., nontrivial multi-step KDD scenarios, rather than just individual data mining operations. The goal of the workshop was to bring together database and data mining researchers interested in the areas of inductive databases, inductive queries, constraint-based data mining, and data mining query languages.

This workshop followed the previous four successful KDID workshops organized in conjunction with ECML/PKDD: KDID 2002 held in Helsinki, Finland, KDID 2003 held in Cavtat-Dubrovnik, Croatia, KDID 2004 held in Pisa, Italy, and KDID 2005 held in Porto, Portugal. Its scientific program included nine regular presentations and two short ones, as well as an invited talk by Kiri L. Wagstaff (Jet Propulsion Laboratory, California Institute of Technology, USA). This volume bundles all papers presented at the workshop and, in addition, includes three contributions that cover relevant research presented at other venues. We also include an article by one of the editors (SD) that attempts to unify existing research in the area and outline directions for further research towards a general framework for data mining.

We wish to thank the invited speaker, all the authors of submitted papers, the program committee members and additional reviewers, and the ECML/PKDD organization committee. KDID 2006 was supported by the European project IQ (“Inductive Queries for Mining Patterns and Models”, IST FET FP6-516169).

July 2007

Sašo Džeroski
Jan Struyf

Organization

Program Chairs

Sašo Džeroski	Department of Knowledge Technologies Jožef Stefan Institute Jamova 39, 1000 Ljubljana, Slovenia saso.dzeroski@ijs.si http://www-ai.ijs.si/SasoDzeroski/
Jan Struyf	Department of Computer Science Katholieke Universiteit Leuven Celestijnenlaan 200A, 3001 Leuven, Belgium jan.struyf@cs.kuleuven.be http://www.cs.kuleuven.be/~jan/

Program Committee

Hiroki Arimura, Hokkaido University, Japan
Hendrik Blockeel, Katholieke Universiteit Leuven, Belgium
Francesco Bonchi, ISTI-C.N.R., Italy
Jean-François Boulicaut, INSA Lyon, France
Toon Calders, University of Antwerp, Belgium
Luc De Raedt, Katholieke Universiteit Leuven, Belgium
Minos N. Garofalakis, Intel Research Berkeley, USA
Fosca Giannotti, ISTI-C.N.R., Italy
Bart Goethals, University of Antwerp, Belgium
Jiawei Han, University Illinois at Urbana-Champaign, USA
Ross D. King, University of Wales, Aberystwyth, UK
Giuseppe Manco, ICAR-C.N.R., Italy
Rosa Meo, University of Turin, Italy
Ryszard S. Michalski, George Mason University, USA
Taneli Mielikäinen, University of Helsinki, Finland
Shinichi Morishita, University of Tokyo, Japan
Siegfried Nijssen, Katholieke Universiteit Leuven, Belgium
Céline Robardet, INSA Lyon, France
Arno Siebes, Utrecht University, The Netherlands
Takashi Washio, Osaka University, Japan
Philip S. Yu, IBM Thomas J. Watson, USA
Mohammed Zaki, Rensselaer Polytechnic Institute, USA
Carlo Zaniolo, UCLA, USA

Additional Reviewers

Annalisa Appice
Marko Bohanec
Emma L. Byrne
Hong Cheng
Amanda Clare

Francesco Folino
Gemma Garriga
Kenneth A. Kaufman
Elio Masciari
Riccardo Ortale

Jimeng Sun
Janusz Wojtusiak

Table of Contents

Invited Talk

Value, Cost, and Sharing: Open Issues in Constrained Clustering	1
<i>Kiri L. Wagstaff</i>	

Contributed Papers

Mining Bi-sets in Numerical Data	11
<i>Jérémy Besson, Céline Robardet, Luc De Raedt, and Jean-François Boulicaut</i>	
Extending the Soft Constraint Based Mining Paradigm	24
<i>Stefano Bistarelli and Francesco Bonchi</i>	
On Interactive Pattern Mining from Relational Databases	42
<i>Francesco Bonchi, Fosca Giannotti, Claudio Lucchese, Salvatore Orlando, Raffaele Perego, and Roberto Trasarti</i>	
Analysis of Time Series Data with Predictive Clustering Trees	63
<i>Sašo Džeroski, Valentin Gjorgjioski, Ivica Slavkov, and Jan Struyf</i>	
Integrating Decision Tree Learning into Inductive Databases	81
<i>Elisa Fromont, Hendrik Blockeel, and Jan Struyf</i>	
Using a Reinforced Concept Lattice to Incrementally Mine Association Rules from Closed Itemsets	97
<i>Arianna Gallo and Rosa Meo</i>	
An Integrated Multi-task Inductive Database VINLEN: Initial Implementation and Early Results	116
<i>Kenneth A. Kaufman, Ryszard S. Michalski, Jarosław Pietrzykowski, and Janusz Wojtusiak</i>	
Beam Search Induction and Similarity Constraints for Predictive Clustering Trees	134
<i>Dragi Kocev, Jan Struyf, and Sašo Džeroski</i>	
Frequent Pattern Mining and Knowledge Indexing Based on Zero-Suppressed BDDs	152
<i>Shin-ichi Minato and Hiroki Arimura</i>	
Extracting Trees of Quantitative Serial Episodes	170
<i>Mirco Nanni and Christophe Rigotti</i>	

IQL: A Proposal for an Inductive Query Language	189
<i>Siegfried Nijssen and Luc De Raedt</i>	
Mining Correct Properties in Incomplete Databases	208
<i>François Rioult and Bruno Crémilleux</i>	
Efficient Mining Under Rich Constraints Derived from Various Datasets	223
<i>Arnaud Soulet, Jiří Kléma, and Bruno Crémilleux</i>	
Three Strategies for Concurrent Processing of Frequent Itemset Queries Using FP-Growth	240
<i>Marek Wojciechowski, Krzysztof Galecki, and Krzysztof Gawronek</i>	
 Discussion Paper	
Towards a General Framework for Data Mining	259
<i>Sašo Džeroski</i>	
Author Index	301