Principles of Data Mining and Knowledge Discovery

4th European Conference, PKDD 2000
Lyon, France, September 13-16, 2000
Proceedings
Table of Contents

Session 1A – Towards Broader Foundations

Multi-relational Data Mining, Using UML for ILP ........................................ 1
  Arno J. Knobbe, Arno Siebes, Hendrik Blockeel, and
  Daniël Van Der Wallen

An Apriori-Based Algorithm for Mining Frequent Substructures from
Graph Data ........................................................................................................... 13
  Akihiro Inokuchi, Takashi Washio, and Hiroshi Motoda

Basis of a Fuzzy Knowledge Discovery System ............................................. 24
  Maurice Bernadet

Session 1B – Rules and Trees

Confirmation Rule Sets .................................................................................... 34
  Dragan Gamberger and Nada Lavrač

Contribution of Dataset Reduction Techniques to Tree-Simplification and
Knowledge Discovery ....................................................................................... 44
  Marc Sebban and Richard Nock

Combining Multiple Models with Meta Decision Trees ............................... 54
  Ljupčo Todorovski and Sašo Džeroski

Session 2A – Databases and Reward-Based Learning

Materialized Data Mining Views .................................................................... 65
  Tadeusz Morzy, Marek Wojciechowski, and Maciej Zakrzewicz

Approximation of Frequency Queries by Means of Free-Sets ..................... 75
  Jean-François Boulicaut, Arthur Bykowsk, and Christophe Rigotti

Application of Reinforcement Learning to Electrical Power System
Closed-Loop Emergency Control ........................................................................ 86
  Christophe Druet, Damien Ernst, and Louis Wehenkel

Efficient Score-Based Learning of Equivalence Classes of Bayesian
Networks ............................................................................................................. 96
  Paul Munteanu and Denis Cau
Session 2B — Classification

Quantifying the Resilience of Inductive Classification Algorithms ............. 106
  Melanie Hilario and Alexandros Kalousis

Bagging and Boosting with Dynamic Integration of Classifiers ................. 116
  Alexey Tsymbal and Seppo Puuronen

Zoomed Ranking: Selection of Classification Algorithms Based on Relevant
  Performance Information ....................................................... 126
  Carlos Soares and Pavel B. Brazdil

Some Enhancements of Decision Tree Bagging .................................... 136
  Pierre Geurts

Session 3A — Association Rules and Exceptions

Relative Unsupervised Discretization for Association Rule Mining ............ 148
  Marcus-Christopher Ludl and Gerhard Widmer

Mining Association Rules: Deriving a Superior Algorithm by Analyzing
  Today's Approaches .............................................................. 159
  Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh

Unified Algorithm for Undirected Discovery of Exception Rules .............. 169
  Einoshin Suzuki and Jan Žytkow

Sampling Strategies for Targeting Rare Groups from a Bank Customer
  Database .................................................................................. 181
  Jean-Hugues Chauchat, Ricco Rakotomalala, and Didier Robert

Session 3B — Instance-Based Discovery

Instance-Based Classification by Emerging Patterns ............................... 191
  Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao

Context-Based Similarity Measures for Categorical Databases .................. 201
  Gautam Das and Heikki Mannila

A Mixed Similarity Measure in Near-Linear Computational Complexity
  for Distance-Based Methods ...................................................... 211
  Ngoc Binh Nguyen and Tu Bao Ho

Fast Feature Selection Using Partial Correlation for Multi-valued
  Attributes ............................................................................... 221
  Stéphane Lallich and Ricco Rakotomalala
Session 4A — Clustering and Classification

Fast Hierarchical Clustering Based on Compressed Data and OPTICS ........ 232
   Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander

Accurate Recasting of Parameter Estimation Algorithms Using Sufficient
Statistics for Efficient Parallel Speed-Up: Demonstrated for Center-Based
Data Clustering Algorithms ....................................................... 243
   Bin Zhang, Meichun Hsu, and George Forman

Predictive Performance of Weighted Relative Accuracy ...................... 255
   Ljupčo Todorovski, Peter Flach, and Nada Lavrač

Quality Scheme Assessment in the Clustering Process ...................... 265
   Maria Halkidi, M. Vazirgiannis, and Y. Batistakis

Session 5A — Time Series

Algorithm for Matching Sets of Time Series .................................. 277
   Iztok Savnik, Georg Lausen, Hans-Peter Kahle, Heinrich Spiecker, and
   Sebastian Hein

MSTS: A System for Mining Sets of Time Series ................................. 289
   Georg Lausen, Iztok Savnik, and Aldar Dougarjapov

Learning First Order Logic Time Series Classifiers: Rules and Boosting . . 299
   Juan J. Rodríguez, Carlos J. Alonso, and Henrik Boström

Posters

Learning Right Sized Belief Networks by Means of a Hybrid Methodology . 309
   Sylvia Acid and Luis M. De Campos

Algorithms for Mining Share Frequent Itemsets Containing Infrequent
Subsets ................................................................. 316
   Brock Barber and Howard J. Hamilton

Discovering Task Neighbourhoods through Landmark Learning
Performances ............................................................... 325
   Hilan Bensusan and Christophe Giraud-Carrier

Induction of Multivariate Decision Trees by Using Dipolar Criteria .......... 331
   Leon Bobrowski and Marek Krętowski

Inductive Logic Programming in Clementine .................................. 337
   Sam Brewer and Tom Khabaza

A Genetic Algorithm-Based Solution for the Problem of Small Disjuncts .. 345
   Deborah R. Carvalho and Alex A. Freitas
Clustering Large, Multi-level Data Sets: An Approach Based on Kohonen Self Organizing Maps.................................................. 353
   Antonio Ciampi and Yves Lechevallier

Trees and Induction Graphs for Multivariate Response.................. 359
   Antonio Ciampi, Djamel A. Zighed, and Jérémy Clech

CEM - Visualisation and Discovery in Email ............................... 367
   Richard Cole, Peter Eklund and Gerd Stumme

Image Access and Data Mining: An Approach ............................... 375
   Chabane Djeraba

Decision Tree Toolkit: A Component-Based Library of Decision Tree Algorithms .................................................. 381
   Nikos Drossos, Athanasios Papagelis, and Dimitris Kalles

Determination of Screening Descriptors for Chemical Reaction Databases . 388
   Laurent Dury, Laurence Leherte, and Daniel P. Vercauteren

Prior Knowledge in Economic Applications of Data Mining............... 395
   A.J. Feelders

Temporal Machine Learning for Switching Control ........................ 401
   Pierre Geurts and Louis Wehenkel

Improving Dissimilarity Functions with Domain Knowledge, Applications with IKBS System ........................................... 409
   David Grosser, Jean Diatta, and Noël Conruyt

Mining Weighted Association Rules for Fuzzy Quantitative Items ........ 416
   Attila Gyenesei

Centroid-Based Document Classification: Analysis and Experimental Results .................................................. 424
   Eui-Hong (Sam) Han and George Karypis

Applying Objective Interestingness Measures in Data Mining Systems .... 432
   Robert J. Hilderman and Howard J. Hamilton

Observational Logic Integrates Data Mining Based on Statistics and Neural Networks ........................................... 440
   Martin Holeňa

Supporting Discovery in Medicine by Association Rule Mining of Bibliographic Databases ........................................... 446
   Dimitar Hristovski, Sašo Džeroski, Borut Peterlin, and Anamarija Rozic-Hristovski
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collective Principal Component Analysis from Distributed, Heterogeneous Data</td>
<td>452</td>
</tr>
<tr>
<td>Hillol Kargupta, Weiyun Huang, Krishnamoorthy Sivakumar, Byung-Hoon Park, and Shuren Wang</td>
<td></td>
</tr>
<tr>
<td>Hierarchical Document Clustering Based on Tolerance Rough Set Model</td>
<td>458</td>
</tr>
<tr>
<td>Saori Kawasaki, Ngoc Binh Nguyen, and Tu Bao Ho</td>
<td></td>
</tr>
<tr>
<td>Application of Data-Mining and Knowledge Discovery in Automotive Data Engineering</td>
<td>464</td>
</tr>
<tr>
<td>Jörg Keller, Valerij Bauer, and Wojciech Kwedlo</td>
<td></td>
</tr>
<tr>
<td>Towards Knowledge Discovery from cDNA Microarray Gene Expression Data</td>
<td>470</td>
</tr>
<tr>
<td>Jan Komorowski, Torger R. Hvidsten, Tor-Kristian Jenssen, Dyre Tjeldvoll, Eivind Hovig, Arne K. Sanvik, and Astrid Lægreid</td>
<td></td>
</tr>
<tr>
<td>Mining with Cover and Extension Operators</td>
<td>476</td>
</tr>
<tr>
<td>Marzena Kryszkiewicz</td>
<td></td>
</tr>
<tr>
<td>A User-Driven Process for Mining Association Rules</td>
<td>483</td>
</tr>
<tr>
<td>Pascale Kuntz, Fabrice Guillet, Rémi Lehn, and Henri Briand</td>
<td></td>
</tr>
<tr>
<td>Learning from Labeled and Unlabeled Documents: A Comparative Study on Semi-Supervised Text Classification</td>
<td>490</td>
</tr>
<tr>
<td>Carsten Lanquillon</td>
<td></td>
</tr>
<tr>
<td>Schema Mining: Finding Structural Regularity among Semistructured Data</td>
<td>498</td>
</tr>
<tr>
<td>P.A. Laur, F. Masseglia, and P. Poncelet</td>
<td></td>
</tr>
<tr>
<td>Improving an Association Rule Based Classifier</td>
<td>504</td>
</tr>
<tr>
<td>Bing Liu, Yiming Ma, and Ching Kian Wong</td>
<td></td>
</tr>
<tr>
<td>Discovery of Generalized Association Rules with Multiple Minimum Supports</td>
<td>510</td>
</tr>
<tr>
<td>Chung-Leung Lui and Fu-Lai Chung</td>
<td></td>
</tr>
<tr>
<td>Learning Dynamic Bayesian Belief Networks Using Conditional Phase-Type Distributions</td>
<td>516</td>
</tr>
<tr>
<td>Adele Marshall, Sally McClean, Mary Shopcott, and Peter Millard</td>
<td></td>
</tr>
<tr>
<td>Discovering Differences in Patients with Uveitis through Typical Testors by Class</td>
<td>524</td>
</tr>
<tr>
<td>José F. Martínez-Trinidad, Miriam Velasco-Sánchez, and Edgar E. Contreras-Aravelo</td>
<td></td>
</tr>
</tbody>
</table>
Web Usage Mining: How to Efficiently Manage New Transactions and New Clients ................................................................. 530
  F. Masseglia, P. Poncelet, and M. Teisseire

Mining Relational Databases ........................................................................ 536
  Frédéric Moal, Teddy Turmeaux, and Christel Vrain

Interestingness in Attribute-Oriented Induction (AOI):
Multiple-Level Rule Generation .................................................................. 542
  Maybin K. Muyeba and John A. Keane

Discovery of Characteristic Subgraph Patterns Using Relative Indexing and the Cascade Model ........................................ 550
  Takashi Okada and Mayumi Oyama

Transparency and Predictive Power: Explaining Complex Classification Models ................................................................. 558
  Gerhard Paass and Jörg Kindermann

Clustering Distributed Homogeneous Datasets ........................................... 566
  Srinivasan Parthasarathy and Mitsunori Ogihara

Empirical Evaluation of Feature Subset Selection Based on a Real-World Data Set ................................................................. 575
  Petra Perner and Chid Apte

Discovery of Ambiguous Patterns in Sequences: Application to Bioinformatics ................................................................. 581
  Gerard Ramstein, Pascal Bunelle, and Yannick Jacques

Action-Rules: How to Increase Profit of a Company .................................. 587
  Zbigniew W. Ras and Alicja Wieczorkowska

Aggregation and Association in Cross Tables ......................................... 593
  Gilbert Ritschard and Nicolas Nicoloyannis

An Experimental Study of Partition Quality Indices in Clustering ............. 599
  Céline Robardet, Fabien Feschet, and Nicolas Nicoloyannis

Expert Constrained Clustering: A Symbolic Approach ............................ 605
  Fabrice Rossi and Frédéric Vautrain

An Application of Association Rules Discovery to Geographic Information Systems .......................................................... 613
  Ansaf Salleb and Christel Vrain

Generalized Entropy and Projection Clustering of Categorical Data .......... 619
  Dan A. Simovici, Dana Cristofor, and Laurentiu Cristofor
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supporting Case Acquisition and Labelling in the Context of Web Mining</td>
<td>626</td>
</tr>
<tr>
<td>Vojtěch Svátek and Martin Kavalec</td>
<td></td>
</tr>
<tr>
<td>Indirect Association: Mining Higher Order Dependencies in Data</td>
<td>632</td>
</tr>
<tr>
<td>Pang-Ning Tan, Vipin Kumar, and Jaideep Srivastava</td>
<td></td>
</tr>
<tr>
<td>Discovering Association Rules in Large, Dense Databases</td>
<td>638</td>
</tr>
<tr>
<td>Tudor Teusan, Gilles Nachouki, Henri Briand, and Jacques Philippe</td>
<td></td>
</tr>
<tr>
<td>Providing Advice to Website Designers Towards Effective Websites</td>
<td>646</td>
</tr>
<tr>
<td>Peter Tselios, Agapios Platis, and George Vouros</td>
<td></td>
</tr>
<tr>
<td>Clinical Knowledge Discovery in Hospital Information Systems: Two Case Studies</td>
<td>652</td>
</tr>
<tr>
<td>Shusaku Tsumoto</td>
<td></td>
</tr>
<tr>
<td>Knowledge Discovery Using Least Squares Support Vector Machine</td>
<td>657</td>
</tr>
<tr>
<td>Classifiers: A Direct Marketing Case</td>
<td></td>
</tr>
<tr>
<td>Stijn Viaene, B. Baesens, T. Van Gestel, J.A.K. Suykens,</td>
<td></td>
</tr>
<tr>
<td>D. Van Den Poel, J. Vanthienen, D. De Moor, and G. Dedene</td>
<td></td>
</tr>
<tr>
<td>Lightweight Document Clustering</td>
<td>665</td>
</tr>
<tr>
<td>Sholom M. Weiss, Brian F. White, and Chidanand V. Apte</td>
<td></td>
</tr>
<tr>
<td>Automatic Category Structure Generation and Categorization of Chinese Text Documents</td>
<td>673</td>
</tr>
<tr>
<td>Hsin-Chang Yang and Chung-Hong Lee</td>
<td></td>
</tr>
<tr>
<td>Mining Generalized Multiple-Level Association Rules</td>
<td>679</td>
</tr>
<tr>
<td>Show-Jane Yen</td>
<td></td>
</tr>
<tr>
<td>An Efficient Approach to Discovering Sequential Patterns in Large Databases</td>
<td>685</td>
</tr>
<tr>
<td>Show-Jane Yen and Chung-Wen Cho</td>
<td></td>
</tr>
<tr>
<td>Using Background Knowledge as a Bias to Control the Rule Discovery</td>
<td>691</td>
</tr>
<tr>
<td>Ning Zhong, Juzhen Dong, and Setsuo Ohsuga</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>699</td>
</tr>
</tbody>
</table>