Intelligent Data Engineering and Automated Learning–IDEAL 2002

Third International Conference
Manchester, UK, August 12-14, 2002
Proceedings
Preface

Data analysis and engineering, and associated learning paradigms, are playing increasingly important roles in an increasing number of application fields. Developments and specialities will benefit many scientific and engineering domains from knowledge/information discovery, data mining and analysis, agents and Internet applications to financial management and bio-informatics.

Intelligent Data Engineering and Automated Learning (IDEAL) is a biennial conference dedicated to emerging and challenging topics in intelligent data analysis and engineering and associated learning paradigms. Following the highly successful IDEAL’98 and IDEAL’00 conferences, both held in Hong Kong, the third conference in this series, IDEAL’02, attracted researchers, scientists and practitioners from all over the world, who came together in Manchester to present their findings and theories, to exchange ideas, and to share their successes. IDEAL has proven to be an ideal forum for revealing and developing the latest theoretical advances and practical applications in intelligent data engineering and automated learning. It is becoming a major international and interdisciplinary event. The themes of IDEAL’02 are Document Analysis and Management, Data Mining, Financial Engineering, Agent Technologies, and Bio-informatics. Over 150 papers were submitted to the conference and they were reviewed by the Program Committee and the additional reviewers. We finally selected about 80 high-quality papers. A special session on Autonomous Mining also contributed a number of excellent papers.

We would like to thank the International Advisory Committee for their guidance and advice, and the Program Committee and additional reviewers for their efficient reviewing of the contributed papers and their helpful comments for the authors. The Asia and America Liaisons did also an excellent job in publicizing the event. We would also like to express our gratitude to the IEEE Neural Networks Society, the UK Engineering and Physical Sciences Research Council (EPSRC), the publisher, Springer-Verlag, and the Manchester Conference Centre for their support throughout.

University of Manchester Institute of Science and Technology (UMIST) Manchester, UK

June 2002

Hujun Yin
Nigel M. Allinson
Richard Freeman
John Keane
Simon Hubbard
Organization

General Co-Chairs

Hujun Yin  UMIST, UK
Nigel Allinson  UMIST, UK
Lei Xu  Chinese University of Hong Kong

International Advisory Committee

Lei Xu (Chair)  Chinese University of Hong Kong
Yaser Abu-Mostafa  CALTECH, USA
Shun-ichi Amari  RIKEN, Japan
Michael Dempster  University of Cambridge, UK
Nick Jennings  University of Southampton, UK
Erkki Oja  Helsinki University of Technology, Finland
Lalit M. Patnaik  Indian Institute of Science, India
Burkhard Rost  Columbia University, USA

Organizing Committee

Hujun Yin (Chair)  UMIST, UK
Nigel Allinson  UMIST, UK
Richard Freeman  UMIST, UK
Simon Hubbard  UMIST, UK
John Keane  UMIST, UK

Asia Liaison

Yiu-ming Cheung  Hong Kong Baptist University, Hong Kong

America Liaison

Malik Magdon-Ismail  Rensselaer Polytechnic Institute, USA

Special Session Organizers

Yiu-ming Cheung  Hong Kong Baptist University, Hong Kong
Jiming Liu  Hong Kong Baptist University, Hong Kong
Program Committee

Nigel Allinson (Chair)  UMIST, UK
Jim Austin  University of York, UK
Hamid Bolouri  University of Hertfordshire, UK
Max Bramer  University of Portsmouth, UK
Laiwan Chan  Chinese University of Hong Kong
Tom Downs  University of Queensland, Australia
Colin Fyfe  University of Paisley, UK
Joydeep Ghosh  University of Texas, USA
Tony Holden  University of Cambridge, UK
Simon Hubbard  UMIST, UK
David Jones  University College London (UCL), UK
Samuel Kaski  Helsinki University of Technology, Finland
John Keane  UMIST, UK
Martin Kersten  CWI Amsterdam, The Netherlands
Irwin King  Chinese University of Hong Kong
Chris Kirkham  AXEON Ltd., UK
Jimmy Lee  Chinese University of Hong Kong
Kwong S. Leung  Chinese University of Hong Kong
Malik Magdon-Ismail  Rensselaer Polytechnic Institute, USA
Luc Moreau  University of Southampton, UK
Jose Principe  University of Florida, USA
Omer Rana  University of Wales, Cardiff, UK
Vic Rayward-Smith  University of East Anglia, UK
Jennie Si  Arizona State University, USA
Ben Stapley  UMIST, UK
Atsuhiro Takasu  National Institute of Informatics, Japan
Marc van Hulle  K. U. Leuven, Belgium
Lipo Wang  Nanyan Technological University, Singapore
Olaf Wolkenhauer  UMIST, UK
Andy Wright  BAE Systems, UK
Xin Yao  University of Birmingham, UK
Xinfeng Ye  University of Auckland, New Zealand
Hujun Yin  UMIST, UK
Hans-Georg Zimmermann  Siemens, Germany
Additional Reviewers

Sophia Ananiadou | Salford University, UK
Zuhair Bandar | Manchester Metropolitan University, UK
Songcan Chen | Nanjing University of Aeronautics and Astronautics, China
Keeley Crockett | Manchester Metropolitan University, UK
Christie Ezeife | University of Windsor, Canada
Richard Freeman | UMIST, UK
Jonathan Gabbai | UMIST/BAE Systems, UK
Ann Gledson | UMIST/Premier Systems Technology, UK
Cefn Hoile | British Telecommunications, UK
Huosheng Hu | Essex University, UK
Yoo-Shin Kim | Pusan National University, Korea
Paulo Lisboa | Liverpool John Moores University, UK
Yuchang Lu | Tsinghua University, China
Farid Meziane | Salford University, UK
Emanuela Moreale | Open University, UK
Andy Nisbet | Trinity College Dublin, Ireland
Ilias Petrounias | UMIST, UK
Ben Russell | UMIST/Premier Systems Technology, UK
Jeevandra Sivarajah | UMIST, UK
Goran Trajkvski | Towson University/West Virginia University, USA

Wenjia Wang | Bradford University, UK
Zhen Rong Yang | Exeter University, UK
Qingfu Zhang | Essex University, UK
# Table of Contents

## Data Mining

Mining Frequent Sequential Patterns under a Similarity Constraint . . . . . 1  
*Matthieu Capelle, Cyrille Masson, Jean-François Boulicaut*

Pre-pruning Classification Trees to Reduce Overfitting in Noisy Domains .................................................. 7  
*Max Bramer*

Data Mining for Fuzzy Decision Tree Structure with a Genetic Program . . 13  
*James F. Smith III*

Co-evolutionary Data Mining to Discover Rules for Fuzzy Resource Management ................................................................. 19  
*James F. Smith III*

Discovering Temporal Rules from Temporally Ordered Data ................. 25  
*Kamran Karimi, Howard J. Hamilton*

Automated Personalisation of Internet Users Using Self-Organising Maps ................................................................. 31  
*Yrjö Hiltunen, Mika Lappalainen*

Data Abstractions for Numerical Attributes in Data Mining .................. 35  
*Masaaki Narita, Makoto Haraguchi, Yoshiaki Okubo*

Calculating Aggregates with Range-Encoded Bit-Sliced Index ............... 43  
*Kashif Bhutta*

T3: A Classification Algorithm for Data Mining ................................ 50  
*Christos Tjortjis, John Keane*

A Hierarchical Model to Support Kansei Mining Process .................... 56  
*Tomofumi Hayashi, Akio Sato, Nadia Berthouze*

Evolving SQL Queries for Data Mining ...................................... 62  
*Majid Salim, Xin Yao*

Indexing and Mining of the Local Patterns in Sequence Database .......... 68  
*Xiaoming Jin, Likun Wang, Yuchang Lu, Chunyi Shi*
# Knowledge Engineering

- A Knowledge Discovery by Fuzzy Rule Based Hopfield Network
  
  Thanakorn Sornkaew, Yasuo Yamashita
  
  Page: 74

- Fusing Partially Inconsistent Expert and Learnt Knowledge in Uncertain Hierarchies
  
  Jonathan Rossiter
  
  Page: 80

- Organisational Information Management and Knowledge Discovery in Email within Mailing Lists
  
  Emanuela Moreale, Stuart Watt
  
  Page: 87

- Design of Multi-drilling Gear Machines by Knowledge Processing and Machine Simulation
  
  G. Klene, A. Grauel, H.J. Convey, A.J. Hartley
  
  Page: 93

# Text and Document Processing

- Classification of Email Queries by Topic: Approach Based on Hierarchically Structured Subject Domain
  
  Anna V. Zhdanova, Denis V. Shishkin
  
  Page: 99

- A Knowledge-Based Information Extraction System for Semi-structured Labeled Documents
  
  Jaeyoung Yang, Heekuck Oh, Kyung-Goo Doh, Joongmin Choi
  
  Page: 105

- Measuring Semantic Similarity Between Words Using Lexical Knowledge and Neural Networks
  
  Yuhua Li, Zuhair Bandar, David Mclean
  
  Page: 111

- Extraction of Hidden Semantics from Web Pages
  
  Vincenza Carchiolo, Alessandro Longheu, Michele Malgeri
  
  Page: 117

- Self-Organising Maps for Hierarchical Tree View Document Clustering Using Contextual Information
  
  Richard Freeman, Hujun Yin
  
  Page: 123

- Schema Discovery of the Semi-structured and Hierarchical Data
  
  Jianwen He
  
  Page: 129

- RSTIndex: Indexing and Retrieving Web Document Using Computational and Linguistic Techniques
  
  Farhi Marir, Kamel Houam
  
  Page: 135

- A Case-Based Recognition of Semantic Structures in HTML Documents (An Automated Transformation from HTML to XML)
  
  Masayuki Umehara, Koji Iwanuma, Hidetomo Nabeshima
  
  Page: 141
Expeditious XML Processing ........................................ 148  
   Kelvin Yeow, R. Nigel Horspool, Michael R. Levy

Document Clustering Using the 1 + 1 Dimensional Self-Organising Map . . 154  
   Ben Russell, Hujun Yin, Nigel M. Allinson

Natural Language Processing for Expertise Modelling in E-mail Communication ............................................... 161  
   Sanghee Kim, Wendy Hall, Andy Keane

**Internet Applications**

A Branch and Bound Algorithm for Minimum Cost Network Flow Problem ................................................................. 167  
   Jun Han, Graham McMahon, Stephen Sugden

Study of the Regularity of the Users’ Internet Accesses ................. 173  
   Nicolas Durand, Luigi Lancieri

An Intelligent Mobile Commerce System with Dynamic Contents Builder and Mobile Products Browser ................................. 179  
   Sera Jang, Eunseok Lee

Focused Crawling Using Fictitious Play ........................................ 186  
   Ville Könönen

A User Adaptive Mobile Commerce System with a Middlet Application ... 193  
   Eunseok Lee, Sera Jang

Weight-Vector Based Approach for Product Recommendation in E-commerce ........................................................... 200  
   Bhanu Prasad

The Development of an XML-Based Data Warehouse System ............ 206  
   Shi-Ming Huang, Chun-Hao Su

Identifying Data Sources for Data Warehouses ................................ 213  
   Christian Koncilia, Heinz Pozewaunig

**Agent Technologies**

Coordinating Learning Agents via Utility Assignment ..................... 219  
   Steven Lynden, Omer F. Rana

AGILE: An Agent-Assisted Infrastructure to Support Learning Environments .......................................................... 225  
   Nicoletta Dessì

Multi-agent Fuzzy Logic Resource Manager .................................. 231  
   James F. Smith III
Transactional Multiple Agents ................................. 237
   Xinfeng Ye, John Keane, Guoqing Zhang

An Information Model for a Merchant Trust Agent in
Electronic Commerce ............................................. 243
   Mohd Khairudin Kasiran, Farid Meziane

MASIVE: A Case Study in Multiagent Systems ................. 249
   Goran Trajkovski

Learning Multi-agent Strategies in Multi-stage Collaborative Games ...... 255
   W. Andy Wright

Emergent Specialization in Swarm Systems ..................... 261
   Ling Li, Alcherio Martinoli, Yaser S. Abu-Mostafa

Distributed Mobile Communication Base Station Diagnosis and
Monitoring Using Multi-agents ..................................... 267
   J.Q. Feng, D.P. Buse, Q.H. Wu, J. Fitch

ABBA - Agent Based Beaver Application - Busy Beaver in Swarm .......... 273
   Alessandro Perrone, Gianluigi Ferraris

Centralised and Distributed Organisational Control ............... 279
   J.M.E. Gabbai, W. Andy Wright, Nigel M. Allinson

Special Session on Autonomous Mining

Mining Dependence Structures from Statistical Learning Perspective ..... 285
   Lei Xu

\( k^* \)-Means — A Generalized \( k \)-Means Clustering Algorithm with
Unknown Cluster Number ............................................. 307
   Yiu-ming Cheung

Multiagent SAT (MASSAT): Autonomous Pattern Search in
Constrained Domains ................................................. 318
   Xiaolong Jin, Jiming Liu

A Text Mining Agents Based Architecture for Personal E-mail
Filtering and Management ........................................... 329
   Ning Zhong, Takahisa Matsunaga, Chunnian Liu

Framework of a Multi-agent KDD System .......................... 337
   Ning Zhong, Yasuaki Matsui, Tomohiro Okuno, Chunnian Liu

Financial Engineering

Intraday FX Trading: An Evolutionary Reinforcement Learning Approach . 347
   M.A.H. Dempster, Y.S. Romahi
An Up-Trend Detection Using an Auto-Associative Neural Network: KOSPI 200 Futures .......................................................... 359
   Jinwoo Baek, Sungzoon Cho

Stock Price and Index Forecasting by Arbitrage Pricing Theory-Based Gaussian TFA Learning ........................................... 366
   Kai Chun Chiu, Lei Xu

A Comparative Study on Three MAP Factor Estimate Approaches for NFA372
   Zhiyong Liu, Lei Xu

A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection ...................................................... 378
   Min-Jung Kim, Taek-Soo Kim

A Comparison of Two Techniques for Next-Day Electricity Price Forecasting ........................................................... 384
   Alicia Troncoso Lora, Jesús Riquelme Santos, José Riquelme Santos,
   Antonio Gómez Expósito, José Luís Martínez Ramos

Support Vector Machine Regression for Volatile Stock Market Prediction................................................................. 391
   Haiqin Yang, Laiwan Chan, Irwin King

Complexity Pursuit for Financial Prediction ................................................. 397
   Ying Han, Colin Fyfe

Artificial Intelligence in Portfolio Management ...................... 403
   Man-Chung Chan, Chi-Cheong Wong, W.F. Tse, Bernard K.-S. Cheung,
   Gordon Y.-N. Tang

The Multilevel Classification Problem and a Monotonicity Hint .............. 410
   Malik Magdon-Ismail, Hung-Ching (Justin) Chen,
   Yaser S. Abu-Mostafa

Adaptive Filtering for GARCH Models .................................................. 416
   Paul E. Lynch, Nigel M. Allinson

Bio-Informatics

Application of Self-Organising Maps in Automated Chemical Shift Correction of In Vivo $^1$H MR Spectra ......................... 423
   Juhani Pulkkinen, Mika Lappalainen, Anna-Maija Häkkinen,
   Nina Lundbom, Risto A. Kauppinen, Yrjö Hiltunen

Supervised Learning of Term Similarities ........................................ 429
   Irena Spasić, Goran Nenadić, Kostas Manios, Sophia Ananiadou
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIKMAS: A Knowledge Engineering System for Bioinformatics</td>
<td>435</td>
</tr>
<tr>
<td>Victoria López-Alonso, Lucia Moreno, Guillermo López-Campos, Victor Maojo, Fernando Martín-Sanchez</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Feature Extraction of <em>in vivo</em> Magnetic Resonance Spectra of Brain Tumours Using Independent Component Analysis</td>
<td>441</td>
</tr>
<tr>
<td>C. Ladroue, A.R. Tate, F.A. Howe, J.R. Griffiths</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Rule-Based Framework for Medical Record Validation</td>
<td>447</td>
</tr>
<tr>
<td>K. Supekar, A. Marwadi, Y. Lee, D. Medhi</td>
<td></td>
</tr>
</tbody>
</table>

### Learning Systems

Classification Learning by Decomposition of Numerical Datasets                                 | 454  |
| Grace J. Hwang, Chun-Chan Tung                                                               |      |
| Combining Feature Selection with Feature Weighting for k-NN Classifier                        | 461  |
| Yongguang Bao, Xiaoyong Du, Naohiro Ishii                                                    |      |
| Pattern Selection for Support Vector Classifiers                                             | 469  |
| Hyunjung Shin, Sungzoon Cho                                                                   |      |
| Graphical Features Selection Method                                                          | 475  |
| Yuan-chin Ivan Chang, Haoran Hsu, Lin-Yi Chou                                                |      |
| Fuzzy-Neural Inference in Decision Trees                                                     | 481  |
| Keeley Crockett, Zuhair Bandar, James O’Shea                                                 |      |
| Decision Tree Based Clustering                                                               | 487  |
| Dongsuk Yook                                                                                 |      |
| Usage of New Information Estimations for Induction of Fuzzy Decision Trees                    | 493  |
| Vitaly G. Levashenko, Elena N. Zaitseva                                                      |      |
| Genetic Algorithm Based-On the Quantum                                                       | 500  |
| Probability Representation                                                                   |      |
| Bin LI, Zhen-quan Zhuang                                                                     |      |
| A Dynamic Method for Discretization of Continuous Attributes                                 | 506  |
| Grace J. Hwang, Fumin Li                                                                     |      |
| A New Neural Implementation of Exploratory Projection Pursuit                               | 512  |
| Colin Fyfe, Emilio Corchado                                                                  |      |
| A General Framework for a Principled Hierarchical Visualization of Multivariate Data         | 518  |
| Ata Kabán, Peter Tiño, Mark Girolami                                                         |      |
Chinese Character Recognition - Comparison of Classification
Methodologies .......................................................... 524
   Sameer Singh, Adnan Amin, K.C. Sum

Lempel-Ziv Coding in Reinforcement Learning ......................... 531
   Kazunori Iwata, Naohiro Ishii

**Pattern Recognition**

Efficient Face Extraction Using Skin-Color Model and a
Neural Network .......................................................... 538
   Jong-Bae Kim, Chae-Hyun Moon, Hang-Joon Kim

Feature Weights Determining of Pattern Classification by Using a
Rough Genetic Algorithm with Fuzzy Similarity Measure ............ 544
   Shan Ding, Naohiro Ishii

Recursive Form of the Discrete Fourier Transform for
Two-Dimensional Signals ............................................... 551
   Zümray Dokur, Tamer Ölmmez

Viseme Recognition Experiment Using Context Dependent Hidden
Markov Models .......................................................... 557
   Soonkyu Lee, Dongsuk Yook

Stave Extraction for Printed Music Scores ........................... 562
   Hidetoshi Miyao

Scaling-Up Model-Based Clustering Algorithm by Working on
Clustering Features .................................................. 569
   Huidong Jin, Kwong-Sak Leung, Man-Leung Wong

A New Approach to Hierarchically Retrieve MPEG Video ............ 576
   Yang Liu, Huanqiang Zhang, Zhimei Wu

Alpha-Beta Search Revisited ........................................ 582
   Paul Parkins, John A. Keane

Quantifying Relevance of Input Features ........................... 588
   Wenjia Wang

**Author Index** .......................................................... 595