

LNAI 3587

Petra Perner
Atsushi Imiya (Eds.)

Machine Learning and Data Mining in Pattern Recognition

4th International Conference, MLDM 2005
Leipzig, Germany, July 2005
Proceedings



Springer

Series Editors

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

Volume Editors

Petra Pernert

Institute of Computer Vision and Applied Computer Sciences, IBal
Körnerstr 10, 04107 Leipzig, Germany
E-mail: ibaiperner@aol.com

Atsushi Imiya

Chiba University, Department of Information and Image Sciences
1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522, Japan
E-mail: imiya@ics.tj.chiba-u.ac.jp

Library of Congress Control Number: 2005928444

CR Subject Classification (1998): I.2, I.5, I.4, F.4.1, H.3

ISSN 0302-9743

ISBN-10 3-540-26923-1 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-26923-6 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper SPIN: 11510888 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3587

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Preface

We met again in front of the statue of Gottfried Wilhelm von Leibniz in the city of Leipzig. Leibniz, a famous son of Leipzig, planned automatic logical inference using symbolic computation, aimed to collate all human knowledge. Today, artificial intelligence deals with large amounts of data and knowledge and finds new information using machine learning and data mining. Machine learning and data mining are irreplaceable subjects and tools for the theory of pattern recognition and in applications of pattern recognition such as bioinformatics and data retrieval.

This was the fourth edition of MLDM in Pattern Recognition which is the main event of Technical Committee 17 of the International Association for Pattern Recognition; it started out as a workshop and continued as a conference in 2003. Today, there are many international meetings which are titled “machine learning” and “data mining”, whose topics are text mining, knowledge discovery, and applications. This meeting from the first focused on aspects of machine learning and data mining in pattern recognition problems. We planned to reorganize classical and well-established pattern recognition paradigms from the viewpoints of machine learning and data mining. Though it was a challenging program in the late 1990s, the idea has inspired new starting points in pattern recognition and effects in other areas such as cognitive computer vision.

For this edition the Program Committee received 103 submissions from 20 countries. After the peer-review process, we accepted 58 papers for presentation. We deeply thank the members of the Program Committee and the reviewers, who examined some difficult papers from broad areas and applications. We also thank the members of the Institute of Applied Computer Sciences, Leipzig, Germany who ran the conference secretariat. We appreciate the help and understanding of the editorial staff at Springer, and in particular Alfred Hofmann, who supported the publication of these proceedings in the LNAI series.

Last, but not least, we wish to thank all the speakers and participants who contributed to the success of the conference.

This proceedings also includes a special selection of papers from the Industrial Conference on Data Mining, ICDM-Leipzig 2005, which we think are also interesting for the audience of this book. We also thank the members of the Program Committee of ICDM 2005 for their valuable work, and all the speakers who made this event a success.

Leipzig,
July 2005

Petra Perner
Atsushi Imiya

Machine Learning and Data Mining in Pattern Recognition MLDM 2005

Co-chairs

Petra Perner

Institute of Computer Vision and Applied
Computer Sciences, IBaI, Leipzig, Germany
Chiba University, Japan

Atsushi Imiya

Program Committee

Agnar Aamodt

NTNU, Norway

Horst Bunke

University of Bern, Switzerland

Max Bramer

University of Portsmouth, UK

Krzysztof Cios

University of Colorado, USA

John Debenham

University of Technology, Australia

Dragan Gamberger

Rudjer Boskovic Institute, Croatia

Lothar Gierl

University of Rostock, Germany

Howard J. Hamilton

University of Regina, Canada

Thomas S. Huang

University of Illinois, USA

Atsushi Imiya

Chiba University, Japan

Horace Ip

City University, Hong Kong, China

Herbert Jahn

Aero Space Center, Germany

Longin Jan Latecki

Temple University, Philadelphia, USA

Adam Krzyzak

Concordia University, Montreal, Canada

Brian Lovell

University of Queensland, Australia

Ryszard Michalski

George Mason University, USA

Donato Malerba

University of Bari, Italy

Fabio Roli

University of Cagliari, Italy

Gabriella Sanniti di Baja

Istituto di Cibernetica, Italy

Michele Sebag

Ecole Polytechnique, France

Arnold Smeulders

University of Amsterdam, The Netherlands

Patrick Wang

Northeastern University, USA

Harry Wechsler

George Mason University, USA

Sholom Weiss

IBM Yorktown Heights, USA

Additional Reviewers

Jörg Cassens
Diego Federici
Magnus Lie Hetland
Helge Langseth
Pavel Petrovic
Amund Tveit
Tomasz Rybak
Marek Kretowski
Zenon A. Sosnowski
Adam Schenker
Bertrand Le Saux
Christophe Ignier
Andreas Schlapbach
Rainer Schmidt
Jilin Tu
Hanning Zhou
Huazhong Ning
Ming Liu
George Q. Chui
Guoqin Cui
Dennis Lin
Charlie Dagli
Zhenqiu Zhang
Yisong Chen
Hau San Wong
Shuo Li
Lars Kulik
Rolf Lakaemper
Aleksandar Lazarevic

Slobodan Vucetic
Caruso Costantina
Annalisa Appice
Antonio Varlaro
Corrado Loglisci
Michelangelo Ceci
Margherita Berardi
Gian Luca Marcialis
Giorgio Giacinto
Roberto Perdisci
Giorgio Fumera
Luca Didaci
Alessandra Serra
Olivier Teytaud
Gelly Sylvain
Mary Jeremie
Nicolas Bredeche
Samuel Landau
Antoine Cornuejols
Jerome Aze
Mathieu Roche
Thang Viet Pham
Gertjan J. Burghouts
Lin Mei
Matthew Ma
Saharon Rosset
Tong Zhang
Ricardo Vilalta

Industrial Conference on Data Mining ICDM 2005

Co-chairs

Petra Perner

Institute of Computer Vision and Applied
Computer Sciences, IBaI, Leipzig, Germany
ECDM, Gengenbach, Germany

Andrea Ahlemeyer-Stubbe

Program Committee

Klaus-Dieter Althoff

Fraunhofer IESE, Germany

Chid Apte

IBM Yorktown Heights, USA

Isabelle Bichindaritz

University of Washington, USA

Leon Bobrowski

Bialystok Technical University, Poland

Marc Boulle

France Télécom, France

Da Deng

University of Otago, New Zealand

Ron S. Kenett

KPA Ltd., Raanana, Israel

Eric Pauwels

CWI, The Netherlands

Rainer Schmidt

University of Rostock, Germany

Stijn Viaene

KULeuven, Belgium

Rob A. Vingerhoeds

Ecole Nationale d'Ingénieurs de Tarbes, France

Lecture Notes in Artificial Intelligence (LNAI)

- Vol. 3587: P. Perner, A. Imiya (Eds.), *Machine Learning and Data Mining in Pattern Recognition*. XVII, 695 pages. 2005.
- Vol. 3571: L. Godo (Ed.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. XVI, 1028 pages. 2005.
- Vol. 3559: P. Auer, R. Meir (Eds.), *Learning Theory*. XI, 692 pages. 2005.
- Vol. 3554: A. Dey, B. Kokinov, D. Leake, R. Turner (Eds.), *Modeling and Using Context*. XIV, 572 pages. 2005.
- Vol. 3533: M. Ali, F. Esposito (Eds.), *Innovations in Applied Artificial Intelligence*. XX, 858 pages. 2005.
- Vol. 3528: P.S. Szczepaniak, J. Kacprzyk, A. Niewiadomski (Eds.), *Advances in Web Intelligence*. XVII, 513 pages. 2005.
- Vol. 3518: T.B. Ho, D. Cheung, H. Liu (Eds.), *Advances in Knowledge Discovery and Data Mining*. XXI, 864 pages. 2005.
- Vol. 3508: P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, M. Winikoff (Eds.), *Agent-Oriented Information Systems II*. X, 227 pages. 2005.
- Vol. 3505: V. Gorodetsky, J. Liu, V. A. Skormin (Eds.), *Autonomous Intelligent Systems: Agents and Data Mining*. XIII, 303 pages. 2005.
- Vol. 3501: B. Kégl, G. Lapalme (Eds.), *Advances in Artificial Intelligence*. XV, 458 pages. 2005.
- Vol. 3492: P. Blache, E. Stabler, J. Busquets, R. Moot (Eds.), *Logical Aspects of Computational Linguistics*. X, 363 pages. 2005.
- Vol. 3488: M.-S. Hacid, N.V. Murray, Z.W. Raś, S. Tsumoto (Eds.), *Foundations of Intelligent Systems*. XIII, 700 pages. 2005.
- Vol. 3476: J. Leite, A. Omicini, P. Torroni, P. Yolum (Eds.), *Declarative Agent Languages and Technologies II*. XII, 289 pages. 2005.
- Vol. 3464: S.A. Brueckner, G.D.M. Serugendo, A. Karageorgos, R. Nagpal (Eds.), *Engineering Self-Organising Systems*. XIII, 299 pages. 2005.
- Vol. 3452: F. Baader, A. Voronkov (Eds.), *Logic for Programming, Artificial Intelligence, and Reasoning*. XI, 562 pages. 2005.
- Vol. 3446: T. Ishida, L. Gasser, H. Nakashima (Eds.), *Massively Multi-Agent Systems I*. XI, 349 pages. 2005.
- Vol. 3438: H. Christiansen, P.R. Skadhauge, J. Villadsen (Eds.), *Constraint Solving and Language Processing*. VIII, 205 pages. 2005.
- Vol. 3430: S. Tsumoto, T. Yamaguchi, M. Numao, H. Motoda (Eds.), *Active Mining*. XII, 349 pages. 2005.
- Vol. 3419: B. Faltings, A. Petcu, F. Fages, F. Rossi (Eds.), *Constraint Satisfaction and Constraint Logic Programming*. X, 217 pages. 2005.
- Vol. 3416: M. Böhlen, J. Gamper, W. Polasek, M.A. Wimmer (Eds.), *E-Government: Towards Electronic Democracy*. XIII, 311 pages. 2005.
- Vol. 3415: P. Davidsson, B. Logan, K. Takadama (Eds.), *Multi-Agent and Multi-Agent-Based Simulation*. X, 265 pages. 2005.
- Vol. 3403: B. Ganter, R. Godin (Eds.), *Formal Concept Analysis*. XI, 419 pages. 2005.
- Vol. 3398: D.-K. Baik (Ed.), *Systems Modeling and Simulation: Theory and Applications*. XIV, 733 pages. 2005.
- Vol. 3397: T.G. Kim (Ed.), *Artificial Intelligence and Simulation*. XV, 711 pages. 2005.
- Vol. 3396: R.M. van Eijk, M.-P. Huget, F. Dignum (Eds.), *Agent Communication*. X, 261 pages. 2005.
- Vol. 3394: D. Kudenko, D. Kazakov, E. Alonso (Eds.), *Adaptive Agents and Multi-Agent Systems II*. VIII, 313 pages. 2005.
- Vol. 3392: D. Seipel, M. Hanus, U. Geske, O. Bartenstein (Eds.), *Applications of Declarative Programming and Knowledge Management*. X, 309 pages. 2005.
- Vol. 3374: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), *Environments for Multi-Agent Systems*. X, 279 pages. 2005.
- Vol. 3371: M.W. Barley, N. Kasabov (Eds.), *Intelligent Agents and Multi-Agent Systems*. X, 329 pages. 2005.
- Vol. 3369: V.R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi (Eds.), *Law and the Semantic Web*. XII, 249 pages. 2005.
- Vol. 3366: I. Rahwan, P. Moraitis, C. Reed (Eds.), *Argumentation in Multi-Agent Systems*. XII, 263 pages. 2005.
- Vol. 3359: G. Grieser, Y. Tanaka (Eds.), *Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets*. XIV, 257 pages. 2005.
- Vol. 3346: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), *Programming Multi-Agent Systems*. XIV, 249 pages. 2005.
- Vol. 3345: Y. Cai (Ed.), *Ambient Intelligence for Scientific Discovery*. XII, 311 pages. 2005.
- Vol. 3343: C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, T. Barkowsky (Eds.), *Spatial Cognition IV*. XIII, 519 pages. 2005.
- Vol. 3339: G.I. Webb, X. Yu (Eds.), *AI 2004: Advances in Artificial Intelligence*. XXII, 1272 pages. 2004.
- Vol. 3336: D. Karagiannis, U. Reimer (Eds.), *Practical Aspects of Knowledge Management*. X, 523 pages. 2004.

- Vol. 3327: Y. Shi, W. Xu, Z. Chen (Eds.), Data Mining and Knowledge Management. XIII, 263 pages. 2005.
- Vol. 3315: C. Lemaître, C.A. Reyes, J.A. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. XX, 987 pages. 2004.
- Vol. 3303: J.A. López, E. Benfenati, W. Dubitzky (Eds.), Knowledge Exploration in Life Science Informatics. X, 249 pages. 2004.
- Vol. 3301: G. Kern-Isberner, W. Rödder, F. Kulmann (Eds.), Conditionals, Information, and Inference. XII, 219 pages. 2005.
- Vol. 3276: D. Nardi, M. Riedmiller, C. Sammut, J. Santos-Victor (Eds.), RoboCup 2004: Robot Soccer World Cup VIII. XVIII, 678 pages. 2005.
- Vol. 3275: P. Perner (Ed.), Advances in Data Mining. VIII, 173 pages. 2004.
- Vol. 3265: R.E. Frederking, K.B. Taylor (Eds.), Machine Translation: From Real Users to Research. XI, 392 pages. 2004.
- Vol. 3264: G. Paliouras, Y. Sakakibara (Eds.), Grammatical Inference: Algorithms and Applications. XI, 291 pages. 2004.
- Vol. 3259: J. Dix, J. Leite (Eds.), Computational Logic in Multi-Agent Systems. XII, 251 pages. 2004.
- Vol. 3257: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web. XVII, 517 pages. 2004.
- Vol. 3249: B. Buchberger, J.A. Campbell (Eds.), Artificial Intelligence and Symbolic Computation. X, 285 pages. 2004.
- Vol. 3248: K.-Y. Su, J. Tsujii, J.-H. Lee, O.Y. Kwong (Eds.), Natural Language Processing – IJCNLP 2004. XVIII, 817 pages. 2005.
- Vol. 3245: E. Suzuki, S. Arikawa (Eds.), Discovery Science. XIV, 430 pages. 2004.
- Vol. 3244: S. Ben-David, J. Case, A. Maruoka (Eds.), Algorithmic Learning Theory. XIV, 505 pages. 2004.
- Vol. 3238: S. Biundo, T. Frühwirth, G. Palm (Eds.), KI 2004: Advances in Artificial Intelligence. XI, 467 pages. 2004.
- Vol. 3230: J.L. Vicedo, P. Martínez-Barco, R. Muñoz, M. Saiz Noeda (Eds.), Advances in Natural Language Processing. XII, 488 pages. 2004.
- Vol. 3229: J.J. Alferes, J. Leite (Eds.), Logics in Artificial Intelligence. XIV, 744 pages. 2004.
- Vol. 3228: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff (Eds.), Formal Approaches to Agent-Based Systems. VIII, 290 pages. 2004.
- Vol. 3215: M.G. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. LVII, 906 pages. 2004.
- Vol. 3214: M.G. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. LVIII, 1302 pages. 2004.
- Vol. 3213: M.G. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LVIII, 1280 pages. 2004.
- Vol. 3209: B. Berendt, A. Hotho, D. Mladenice, M. van Someren, M. Spiliopoulou, G. Stumme (Eds.), Web Mining: From Web to Semantic Web. IX, 201 pages. 2004.
- Vol. 3206: P. Sojka, I. Kopecek, K. Pala (Eds.), Text, Speech and Dialogue. XIII, 667 pages. 2004.
- Vol. 3202: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Knowledge Discovery in Databases: PKDD 2004. XIX, 560 pages. 2004.
- Vol. 3201: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Machine Learning: ECML 2004. XVIII, 580 pages. 2004.
- Vol. 3194: R. Camacho, R. King, A. Srinivasan (Eds.), Inductive Logic Programming. XI, 361 pages. 2004.
- Vol. 3192: C. Bussler, D. Fensel (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. XIII, 522 pages. 2004.
- Vol. 3191: M. Klusch, S. Ossowski, V. Kashyap, R. Unland (Eds.), Cooperative Information Agents VIII. XI, 303 pages. 2004.
- Vol. 3187: G. Lindemann, J. Denzinger, I.J. Timm, R. Unland (Eds.), Multiagent System Technologies. XIII, 341 pages. 2004.
- Vol. 3176: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Advanced Lectures on Machine Learning. IX, 241 pages. 2004.
- Vol. 3171: A.L.C. Bazzan, S. Labidi (Eds.), Advances in Artificial Intelligence – SBIA 2004. XVII, 548 pages. 2004.
- Vol. 3159: U. Visser, Intelligent Information Integration for the Semantic Web. XIV, 150 pages. 2004.
- Vol. 3157: C. Zhang, H. W. Guesgen, W.K. Yeap (Eds.), PRICAI 2004: Trends in Artificial Intelligence. XX, 1023 pages. 2004.
- Vol. 3155: P. Funk, P.A. González Calero (Eds.), Advances in Case-Based Reasoning. XIII, 822 pages. 2004.
- Vol. 3139: F. Iida, R. Pfeifer, L. Steels, Y. Kuniyoshi (Eds.), Embodied Artificial Intelligence. IX, 331 pages. 2004.
- Vol. 3131: V. Torra, Y. Narukawa (Eds.), Modeling Decisions for Artificial Intelligence. XI, 327 pages. 2004.
- Vol. 3127: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (Eds.), Conceptual Structures at Work. XI, 403 pages. 2004.
- Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), Natural Language Generation. X, 219 pages. 2004.
- Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), Learning Theory. X, 648 pages. 2004.
- Vol. 3097: D. Basin, M. Rusinowitch (Eds.), Automated Reasoning. XII, 493 pages. 2004.
- Vol. 3071: A. Omicini, P. Petta, J. Pitt (Eds.), Engineering Societies in the Agents World. XIII, 409 pages. 2004.
- Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004.
- Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004.
- Vol. 3067: M. Dastani, J. Dix, A. El Fallah-Seghrouchni (Eds.), Programming Multi-Agent Systems. X, 221 pages. 2004.

¥755.20元

Table of Contents

Classification and Model Estimation

On ECOC as Binary Ensemble Classifiers <i>J. Ko, E. Kim</i>	1
Incremental Classification Rules Based on Association Rules Using Formal Concept Analysis <i>Anamika Gupta, Naveen Kumar, Vasudha Bhatnagar</i>	11
Parameter Inference of Cost-Sensitive Boosting Algorithms <i>Yanmin Sun, Andrew K.C. Wong, Yang Wang</i>	21
Finite Mixture Models with Negative Components <i>Baibo Zhang, Changshui Zhang</i>	31
MML-Based Approach for Finite Dirichlet Mixture Estimation and Selection <i>Nizar Bouguila, Djemel Ziou</i>	42
Principles of Multi-kernel Data Mining <i>Vadim Mottl, Olga Krasotkina, Oleg Seredin, Ilya Muchnik</i>	52

Neural Methods

Comparative Analysis of Genetic Algorithm, Simulated Annealing and Cutting Angle Method for Artificial Neural Networks <i>Ranadhir Ghosh, Moumita Ghosh, John Yearwood, Adil Bagirov</i>	62
Determining Regularization Parameters for Derivative Free Neural Learning <i>Ranadhir Ghosh, Moumita Ghosh, John Yearwood, Adil Bagirov</i>	71
A Comprehensible SOM-Based Scoring System <i>Johan Huysmans, Bart Baesens, Jan Vanthienen</i>	80

Subspace Methods

The Convex Subclass Method: Combinatorial Classifier Based on a Family of Convex Sets <i>Ichigaku Takigawa, Mineichi Kudo, Atsuyoshi Nakamura</i>	90
---	----

SSC: Statistical Subspace Clustering
*Laurent Candillier, Isabella Tellier, Fabien Torre,
Olivier Bousquet* 100

Understanding Patterns with Different Subspace Classification
Gero Szepannek, Karsten Luebke, Claus Weihs 110

Clustering: Basics

Using Clustering to Learn Distance Functions for Supervised Similarity
Assessment
*Christoph F. Eick, Alain Rouhana, Abraham Bagherjeiran,
Ricardo Vilalta* 120

Linear Manifold Clustering
Robert Haralick, Rave Harpaz 132

Universal Clustering with Regularization in Probabilistic Space
Vladimir Nikulin, Alex J. Smola 142

Acquisition of Concept Descriptions by Conceptual Clustering
Silke Jänichen, Petra Perner 153

Applications of Clustering

Clustering Large Dynamic Datasets Using Exemplar Points
William Sia, Mihai M. Lazarescu 163

Birds of a Feather Surf Together: Using Clustering Methods to Improve
Navigation Prediction from Internet Log Files
Martin Halvey, Mark T. Keane, Barry Smyth 174

Alarm Clustering for Intrusion Detection Systems in Computer
Networks
Giorgio Giacinto, Roberto Perdisci, Fabio Roli 184

Clustering Document Images Using Graph Summaries
Eugen Barbu, Pierre Héroux, Sébastien Adam, Eric Trupin 194

**Feature Grouping, Discretization, Selection and
Transformation**

Feature Selection Method Using Preferences Aggregation
Gaëlle Legrand, Nicolas Nicoloyannis 203

Ranked Modelling with Feature Selection Based on the <i>CPL</i> Criterion Functions	
<i>Leon Bobrowski</i>	218
A Grouping Method for Categorical Attributes Having Very Large Number of Values	
<i>Marc Boullé</i>	228
Unsupervised Learning of Visual Feature Hierarchies	
<i>Fabien Scalzo, Justus Piater</i>	243
Multivariate Discretization by Recursive Supervised Bipartition of Graph	
<i>Sylvain Ferrandiz, Marc Boullé</i>	253
CorePhrase: Keyphrase Extraction for Document Clustering	
<i>Khaled M. Hammouda, Diego N. Matute, Mohamed S. Kamel</i>	265
A New Multidimensional Feature Transformation for Linear Classifiers and Its Applications	
<i>EunSang Bak</i>	275
 Applications in Medicine	
Comparison of FLDA, MLP and SVM in Diagnosis of Lung Nodule	
<i>Aristófanés Corrêa Silva, Anselmo Cardoso de Paiva, Alexandre Cesar Muniz de Oliveira</i>	285
Diagnosis of Lung Nodule Using Reinforcement Learning and Geometric Measures	
<i>Aristófanés Corrêa Silva, Valdeci Ribeiro da Silva Junior, Areolino de Almeida Neto, Anselmo Cardoso de Paiva</i>	295
Iris Recognition Algorithm Based on Point Covering of High-Dimensional Space and Neural Network	
<i>Wenming Cao, Jianhui Hu, Gang Xiao, Shoujue Wang</i>	305
Automatic Clinical Image Segmentation Using Pathological Modelling, PCA and SVM	
<i>Shuo Li, Thomas Fevens, Adam Krzyżak, Song Li</i>	314
Improved MRI Mining by Integrating Support Vector Machine Priors in the Bayesian Restoration	
<i>D.A. Karras, B.G. Mertzios, D. Graveron-Demilly, D. van Ormondt</i>	325

Prediction of Secondary Protein Structure Content from Primary
Sequence Alone – A Feature Selection Based Approach 334
Lukasz Kurgan, Leila Homaeian

Alternative Clustering by Utilizing Multi-objective Genetic Algorithm
with Linked-List Based Chromosome Encoding 346
Jun Du, Emin Erkan Korkmaz, Reda Alhaji, Ken Barker

Time Series and Sequential Pattern Mining

Embedding Time Series Data for Classification 356
Akira Hayashi, Yuko Mizuhara, Nobuo Suematsu

Analysis of Time Series of Graphs: Prediction of Node Presence by
Means of Decision Tree Learning 366
Horst Bunke, Peter Dickinson, Christophe Irniger, Miro Kraetzl

Disjunctive Sequential Patterns on Single Data Sequence and Its
Anti-monotonicity 376
Kazuhiro Shimizu, Takao Miura

Mining Expressive Temporal Associations from Complex Data 384
Keith A. Pray, Carolina Ruiz

Statistical Supports for Frequent Itemsets on Data Streams 395
*Pierre-Alain Laur, Jean-Emile Symphor, Richard Nock,
Pascal Poncelet*

Mining Images in Computer Vision

Autonomous Vehicle Steering Based on Evaluative Feedback by
Reinforcement Learning 405
Klaus-Dieter Kuhnert, Michael Krödel

Cost Integration in Multi-step Viewpoint Selection for Object
Recognition 415
Christian Derichs, Frank Deinzer, Heinrich Niemann

Support Vector Machine Experiments for Road Recognition in High
Resolution Images 426
J.Y. Lai, A. Sowmya, J. Trinder

An Automatic Face Recognition System in the Near Infrared Spectrum <i>Shuyan Zhao, Rolf-Rainer Grigat</i>	437
--	-----

Mining Images and Texture

Hierarchical Partitions for Content Image Retrieval from Large-Scale Database <i>Dmitry Kinoshenko, Vladimir Mashtalir, Elena Yegorova, Vladimir Vinarsky</i>	445
Optimising the Choice of Colours of an Image Database for Dichromats <i>Vassili Kovalev, Maria Petrou</i>	456
An Approach to Mining Picture Objects Based on Textual Cues <i>Adeoye I. Adegorite, Otman A. Basir, Mohamed S. Kamel, Khaled B. Shaban</i>	466

Mining Motion from Sequence

Activity and Motion Detection Based on Measuring Texture Change <i>Longin Jan Latecki, Roland Mieziako, Dragoljub Pokrajac</i>	476
A New Approach to Human Motion Sequence Recognition with Application to Diving Actions <i>Shiming Xiang, Changshui Zhang, Xiaoping Chen, Naijiang Lu</i>	487
Dominant Plane Detection Using Optical Flow and Independent Component Analysis <i>Naoya Ohnishi, Atsushi Imiya</i>	497

Speech Analysis

Neural Expert Model Applied to Phonemes Recognition <i>Halima Bahi, Mokhtar Sellami</i>	507
An Evidential Reasoning Approach to Weighted Combination of Classifiers for Word Sense Disambiguation <i>Cuong Anh Le, Van-Nam Huynh, Akira Shimazu</i>	516

Aspects of Data Mining

Signature-Based Approach for Intrusion Detection <i>Bon K. Sy</i>	526
--	-----

Discovery of Hidden Correlations in a Local Transaction Database
Based on Differences of Correlations

Tsuyoshi Taniguchi, Makoto Haraguchi, Yoshiaki Okubo 537

An Integrated Approach for Mining Meta-rules

*Feiyue Ye, Jiandong Wang, Shiliang Wu, Huiping Chen,
Tianqiang Huang, Li Tao* 549

Data Mining on Crash Simulation Data

*Annette Kuhlmann, Ralf-Michael Vetter, Christoph Lübbing,
Clemens-August Thole* 558

Text Mining

Pattern Mining Across Domain-Specific Text Collections

Lee Gillam, Khurshid Ahmad 570

Text Classification Using Small Number of Features

Masoud Makrehchi, Mohamed S. Kamel 580

Low-Level Cursive Word Representation Based on Geometric
Decomposition

*Jian-xiong Dong, Adam Krzyżak, Ching Y. Suen,
Dominique Ponson* 590

Special Track: Industrial Applications of Data Mining

Supervised Evaluation of Dataset Partitions: Advantages and Practice

Sylvain Ferrandiz, Marc Boullé 600

Inference on Distributed Data Clustering

Josenildo C. da Silva, Matthias Klusch 610

A Novel Approach of Multilevel Positive and Negative Association
Rule Mining for Spatial Databases

L.K. Sharma, O.P. Vyas, U.S. Tiwary, R. Vyas 620

Mixture Random Effect Model Based Meta-analysis for Medical Data
Mining

Yinglong Xia, Shifeng Weng, Changshui Zhang, Shao Li 630

Semantic Analysis of Association Rules via Item Response Theory

Shinichi Hamano, Masako Sato 641

Temporal Approach to Association Rule Mining Using T-Tree and P-Tree <i>Keshri Verma, O.P. Vyas, Ranjana Vyas</i>	651
Aquaculture Feature Extraction from Satellite Image Using Independent Component Analysis <i>JongGyu Han, KwangHoon Chi, YeonKwang Yeon</i>	660
Modeling the Organoleptic Properties of Matured Wine Distillates <i>S.B. Kotsiantis, G.E. Tsekouras, C. Raptis, P.E. Pintelas</i>	667
Bagging Random Trees for Estimation of Tissue Softness <i>S.B. Kotsiantis, G.E. Tsekouras, P.E. Pintelas</i>	674
Concept Mining for Indexing Medical Literature <i>Isabelle Bichindaritz, Sarada Akkineni</i>	682
Author Index	693

On ECOC as Binary Ensemble Classifiers

J. Ko¹ and E. Kim²

¹ Dept. of Computer Engineering, Kumoh National Institute of Technology,
1, Yangho-dong, Gumi, Gyeongbuk 730-701, Korea
nonezero@kumloh.ac.kr,

² National Computerization Agency,
NCA Bldg 77, Mugyo-dong, Jung-gu, Seoul 100-775, Korea
outframe@nca.go.kr

Abstract. The Error-Correcting Output Codes (ECOC) is a representative approach of the binary ensemble classifiers for solving multi-class problems. There have been so many researches on an output coding method built on an ECOC foundation. In this paper, we revisit representative conventional ECOC methods in an overlapped learning viewpoint. For this purpose, we propose new OPC based output coding methods in the ECOC point of view, and define a new measure to describe their properties. From the experiment on a face recognition domain, we investigate whether a problem complexity is more important than the overlapped learning or an error correction concept.

1 Introduction

The Error-Correcting Output Codes (ECOC) [1] is one of the binary ensemble classifiers for solving multi-class problems. The ECOC has been dominant theoretical foundation in output coding methods [2-6] that decompose a complex multi-class problem into a set of binary problems and then reconstructs the outputs of binary classifiers for each binary problem. The performance of output coding methods depends on base binary classifiers. It needs to revisit the ECOC concept, since the Support Vector Machines (SVM) [7] that can produce a complex nonlinear decision boundary with a good generalization performance is available as a base classifier for output coding methods.

The ECOC has two principals with respect to a codes design in which the codes concern both how to decompose a multi-class problem into several binary ones and how to decide a final decision. One principal is to enlarge the minimum hamming distance of a decomposition matrix. The other is to enlarge the row separability to increase the diversity among binary problems. A high diversity reduces an error-correlation among binary machines [8]. By enlarging the length of codewords [9], we can easily increase the hamming distance of the decomposition matrix at the cost of generating a large number of binary problems. In this circumstance, each class can be learned redundantly in several binary machines, we call it *overlapped learning*. By increasing the error-correction ability through the overlapped learning, we have been