

# Lecture Notes in Artificial Intelligence 2123

Subseries of Lecture Notes in Computer Science

Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

**Springer**

*Berlin*

*Heidelberg*

*New York*

*Barcelona*

*Hong Kong*

*London*

*Milan*

*Paris*

*Singapore*

*Tokyo*

Petra Perner (Ed.)

# Machine Learning and Data Mining in Pattern Recognition

Second International Workshop, MLDM 2001  
Leipzig, Germany, July 25-27, 2001  
Proceedings



Springer

## Series Editors

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

## Volume Editor

Petra Perner  
Institute of Computer Vision and Applied Computer Sciences  
Arno-Nitzsche-Str. 45, 04277 Leipzig, Germany  
E-mail: ibaiperner@aol.com

## Cataloging-in-Publication Data applied for

### Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Machine learning and data mining in pattern recognition : second  
international workshop ; proceedings / MLDM 2001, Leipzig, Germany, July  
25 - 27, 2001. Petra Perner (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ;  
Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001  
(Lecture notes in computer science ; 2123 : Lecture notes in artificial  
intelligence)  
ISBN 3-540-42359-1

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

ISBN 3-540-42359-1 Springer-Verlag 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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York  
a member of BertelsmannSpringer Science+Business Media GmbH  
<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2001  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Steingraber Satztechnik GmbH, Heidelberg  
Printed on acid-free paper SPIN: 10839922 06/3142 5 4 3 2 1 0

# Preface

The papers published in this book were presented at the second workshop on Machine Learning and Data Mining in Pattern Recognition MLDM. Those of you familiar with the first workshop will notice that the ideas related to this topic are spreading to many researchers. We received several excellent papers on subjects ranging from basic research to application oriented research. The subjects are Case-Based Reasoning, Rule Induction, Grammars, Clustering, Data Mining on Multimedia Data, Content-Based Image Retrieval, Statistical and Evolutionary Learning, Neural Networks, and Learning for Handwriting Recognition. The whole spectrum of topics in MLDM was represented at the workshop with emphasis on images, text, and signals, and temporal spatial data. We also took a step in the direction of our decision made at the last TC3 Machine Learning meeting in Barcelona, to introduce the field to researchers outside the computer science or pattern recognition community. We welcomed medics, specialists of data base marketing, and mechanical engineers to our workshop. These researchers reported their experience and the problems they have in applying Data Mining. By sharing their experience with us they give new impulses to our work.

The workshop was organized by the Leipzig Institute of Computer Vision and Applied Computer Sciences. Many thanks to Maria Petrou for co-chairing MLDM 2001 with me.

It is my pleasure to thank the invited speakers for accepting our invitation to give lectures and contribute papers to the proceedings. I would also like to express my appreciation to the reviewers for their precise and highly professional work. I appreciate the help and understanding of the editorial staff at Springer-Verlag, and in particular Alfred Hofmann, who supported the publication of these proceedings in the LNAI series.

Last but not least, I wish to thank all the speakers and participants for their interest in this workshop. I hope you enjoyed the workshop and that you will return to present your new ideas at MLDM 2003.

July 2001

Petra Perner

# IAPR International Workshop on Machine Learning and Data Mining in Pattern Recognition

## Co-chairs

**Petra Perner**

IBaI Leipzig / Germany

**Maria Petrou**

University of Surrey / United Kingdom

Adnan Amin	Univ. of New South Wales, Australia
Bir Bhanu	Univ. of California, Riverside, USA
Terri Caelli	Univ. of Alberta, Canada
Irwin King	Chinese Univ., Hong Kong
Donato Malerba	University of Bari, Italy
Petra Perner	IBaI, Germany
Maria Petrou	University of Surrey, UK
Fabio Roli	University of Cagliari, Italy
Arnold Smeulders	University of Amsterdam, The Netherlands
Sholom Weiss	IBM Yorktown Heights, USA
Ari Visa	Tampere University, Finland

## The Aim of the Workshop

The aim of the workshop was to bring together researchers from all over the world dealing with machine learning and data mining in order to discuss the recent status of the research and to direct further developments. All kinds of application were welcome. Special preference was given to multimedia related applications.

It was the second workshop in a series of workshops dealing with this specific topic. The first workshop was published in *P. Perner and M. Petrou, Machine Learning and Data Mining in Pattern Recognition MLDM99, LNAI 1715, Springer Verlag 1999* and in a special issue of *Pattern Recognition Letters*.

The topics covered include:

- inductive learning including decision trees
- rule induction learning
- conceptual learning
- case-based learning
- statistical learning
- neural net based learning
- organisational learning
- evolutionary learning
- probabilistic information retrieval

Applications include but are not limited to medical, industrial, and biological applications.

Researchers from the machine learning community were invited to present new topics in learning, pertinent to our research field.

# Table of Contents

## Invited Paper

Technology of Text Mining .....	1
<i>A. Visa</i>	
Evaluation of Clinical Relevance of Clinical Laboratory Investigations by Data Mining .....	12
<i>U. Sack and M. Kamprad</i>	

## Case-Based Reasoning and Associative Memory

Temporal Abstractions and Case-Based Reasoning for Medical Course Data: Two Prognostic Applications .....	23
<i>R. Schmidt and L. Gierl</i>	
Are Case-Based Reasoning and Dissimilarity-Based Classification Two Sides of the Same Coin? .....	35
<i>P. Perner</i>	
FAM-Based Fuzzy Inference for Detecting Shot Transitions .....	52
<i>S.-W. Jang, G.-Y. Kim, and H.-I. Choi</i>	

## Rule Induction and Grammars

Rule-Based Ensemble Solutions for Regression .....	62
<i>N. Indurkha and S. M. Weiss</i>	
Learning XML Grammars .....	73
<i>H. Fernau</i>	
First-Order Rule Induction for the Recognition of Morphological Patterns in Topographic Maps .....	88
<i>D. Malerba, F. Esposito, A. Lanza, and F.A. Lisi</i>	

## Clustering and Conceptual Clustering

Concepts Learning with Fuzzy Clustering and Relevance Feedback .....	102
<i>B. Bhanu and A. Dong</i>	
LC: A Conceptual Clustering Algorithm .....	117
<i>J.F. Martínez-Trinidad and G. Sánchez-Díaz</i>	

**Data Mining on Signal, Images, Text  
and Temporal-Spatial Data**

Data Mining Approach Based on Information-Statistical Analysis:  
Application to Temporal-Spatial Data ..... 128  
*B.K. Sy and A.K. Gupta*

A Hybrid Tool for Data Mining in Picture Archiving System ..... 141  
*P. Perner and T. Belikova*

Feature Selection for a Real-World Learning Task ..... 157  
*D. Kollmar and D.H. Hellmann*

Automatic Identification of Diatoms Using Decision Forests ..... 173  
*S. Fischer and H. Bunke*

Validation of Text Clustering Based on Document Contents ..... 184  
*J. Toivonen, A. Visa, T. Vesanen, B. Back, and H. Vanharanta*

**Nonlinear Function Learning  
and Neural Net Based Learning**

Statistical and Neural Approaches for Estimating Parameters  
of a Speckle Model Based on the Nakagami Distribution ..... 196  
*M.P. Wachowiak, R. Smolíková, M.G. Milanova, and A.S. Elmaghraby*

How to Automate Neural Net Based Learning ..... 206  
*R. Linder and S.J. Pöpl*

Nonlinear Function Learning and Classification  
Using Optimal Radial Basis Function Networks ..... 217  
*A. Krzyżak*

**Learning for Handwriting Recognition**

Local Learning Framework for Recognition  
of Lowercase Handwritten Characters ..... 226  
*J.-x. Dong, A. Krzyżak, and C.Y. Suen*

Mirror Image Learning for Handwritten Numeral Recognition ..... 239  
*M. Shi, T. Wakabayashi, W. Ohyama, and F. Kimura*

**Statistical and Evolutionary Learning**

Face Detection by Aggregated Bayesian Network Classifiers ..... 249  
*T.V. Pham, M. Worring, and A.W.M. Smeulders*

Towards Self-Exploring Discriminating Features ..... 263  
*Y. Wu and T.S. Huang*



PCA-Based Model Selection and Fitting for Linear Manifolds . . . . .	278
<i>A. Imiya and H. Ootani</i>	
Statistics of Flow Vectors and Its Application to the Voting Method for the Detection of Flow Fields . . . . .	293
<i>A. Imiya and K. Iwawaki</i>	
On the Use of Pairwise Comparison of Hypotheses in Evolutionary Learning Applied to Learning from Visual Examples . . . . .	307
<i>K. Krawiec</i>	
Featureless Pattern Recognition in an Imaginary Hilbert Space and Its Application to Protein Fold Classification . . . . .	322
<i>V. Mottl, S. Dvoenko, O. Seredin, C. Kulikowski, and I. Muchnik</i>	
<b>Content-Based Image Retrieval</b>	
Adaptive Query Shifting for Content-Based Image Retrieval . . . . .	337
<i>G. Giacinto, F. Roli, and G. Fumera</i>	
Content-Based Similarity Assessment in Multi-segmented Medical Image Data Bases . . . . .	347
<i>G. Potamias</i>	
<b>Author Index</b> . . . . .	363