Petra Perner
Atsushi Imiya (Eds.)

LNAI 3587

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 Perner
Institute of Computer Vision and Applied Computer Sciences, IBaI
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

Springer is a part of Springer Science+Business Media

springeronline.com

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
Atsushi Imiya	Chiba University, Japan

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	Northeasten 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 Irniger
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 Serrau
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
Andrea Ahlemeyer-Stubbe	ECDM, Gengenbach, Germany

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. 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. Mladenic, 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. Tądeusiewicz, 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

Neural Methods

Subspace Methods

Clustering: Basics

Applications of Clustering

Feature Grouping, Discretization, Selection and Transformation

Applications in Medicine

Time Series and Sequential Pattern Mining

Mining Images in Computer Vision

Mining Images and Texture

Mining Motion from Sequence

Speech Analysis

Aspects of Data Mining

Text Mining

Special Track: Industrial Applications of Data Mining

On ECOC as Binary Ensemble Classifiers

J. Ko[1] and E. Kim[2]

[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

P. Perner and A. Imiya (Eds.): MLDM 2005, LNAI 3587, pp. 1 – 10, 2005.