

Lecture Notes in Artificial Intelligence 2777

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Springer

Berlin

Heidelberg

New York

Hong Kong

London

Milan

Paris

Tokyo

Bernhard Schölkopf Manfred K. Warmuth (Eds.)

Learning Theory and Kernel Machines

16th Annual Conference on Learning Theory
and 7th Kernel Workshop, COLT/Kernel 2003
Washington, DC, USA, August 24-27, 2003
Proceedings



Springer

Series Editors

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

Volume Editors

Bernhard Schölkopf
Max Planck Institute for Biological Cybernetics
Empirical Inference Department
Spemannstr. 38, 72076 Tübingen, Germany
E-mail: bs@tuebingen.mpg.de

Manfred K. Warmuth
University of California
Department of Computer Science
227 Baskin Engineering, Santa Cruz, CA 95064, USA
E-mail: manfred@cse.ucsc.edu

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

CR Subject Classification (1998): I.2.6, I.2.3, I.2, F.4.1, F.2, F.1.1

ISSN 0302-9743

ISBN 3-540-40720-0 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 2003
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 10931424 06/3142 5 4 3 2 1 0

Preface

This volume contains papers presented at the joint 16th Annual Conference on Learning Theory (COLT) and the 7th Annual Workshop on Kernel Machines, held in Washington, DC, USA, during August 24–27, 2003. COLT, which recently merged with EuroCOLT, has traditionally been a meeting place for learning theorists. We hope that COLT will benefit from the collocation with the annual workshop on kernel machines, formerly held as a NIPS postconference workshop.

The technical program contained 47 papers selected from 92 submissions. All 47 papers were presented as posters; 22 of the papers were additionally presented as talks. There were also two target areas with invited contributions. In *computational game theory*, a tutorial entitled “Learning Topics in Game-Theoretic Decision Making” was given by Michael Littman, and an invited paper on “A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria” was contributed by Amy Greenwald. In *natural language processing*, a tutorial on “Machine Learning Methods in Natural Language Processing” was presented by Michael Collins, followed by two invited talks, “Learning from Uncertain Data” by Mehryar Mohri and “Learning and Parsing Stochastic Unification-Based Grammars” by Mark Johnson.

In addition to the accepted papers and invited presentations, we solicited short open problems that were reviewed and included in the proceedings. We hope that reviewed open problems might become a new tradition for COLT. Our goal was to select simple signature problems whose solutions are likely to inspire further research. For some of the problems the authors offered monetary rewards. Yoav Freund acted as the open problem area chair. The open problems were presented as posters at the conference.

The Mark Fulk award is presented annually for the best paper contributed by a student. This year’s award was won by Ulrike von Luxburg for the paper “Distance-Based Classification with Lipschitz Functions” (co-authored with Olivier Bousquet). Owing to generous support from the Journal of Machine Learning, we were able to award two second prizes for the best student paper. They went to Gilles Stoltz, “Internal Regret in On-line Portfolio Selection” (co-authored with Gabor Lugosi), and Petra Philips, “Random Subclass Bounds” (co-authored with Shahar Mendelson).

We thank all the people and organizations who helped make this conference successful. We are especially grateful to the program committee for their efforts in evaluating and selecting the papers in this volume:

Kristin Bennett	Rensselaer Polytechnic Institute
Avrim Blum	Carnegie Mellon University
Nicolò Cesa-Bianchi	University of Milan
Nello Cristianini	University of California at Davis
Yoav Freund	Banter Inc.
Michael Kearns	University of Pennsylvania
Efim Kinber	Sacred Heart University, Fairfield, CT
Vladimir Koltchinskii	The University of New Mexico
Yishay Mansour	Tel Aviv University
Rob Schapire	Princeton University

The program committee solicited the help of the following reviewers. We obtained at least two independent reviews for each paper in addition to the reviews and discussions carried out by the program committee:

Naoki Abe	Rusins Freivalds	Wolfgang Merkle
Dana Angluin	Nir Friedman	Jochen Nessel
Peter Auer	Bill Gasarch	Kamal Nigam
Peter Bartlett	Claudio Gentile	Noam Nisan
Satinder Baveja	Sally Goldman	Bill Noble
Jonathan Baxter	Geoff Gordon	Luis Ortiz
Amos Beimel	Thore Graepel	Dmitry Panchenko
Shai Ben-David	Peter Grunwald	Relu Patrascu
Yoshua Bengio	Patrick Haffner	John Platt
Gilles Blanchard	Lisa Hellerstein	Gunnar Rätsch
Stephane Boucheron	Jeff Jackson	Vijay Raghavan
Olivier Bousquet	Thorsten Joachims	Dana Ron
Nader Bshouty	Adam Kalai	Roman Rosipal
Chris Burges	Jaz Kandola	Dale Schuurmans
Colin Campbell	Balazs Kegl	Rocco Servedio
Rafael Carrasco	Jyrki Kivinen	John Shawe-Taylor
John Case	John Lafferty	Yoram Singer
Moses Charikar	Gert Lanckriet	Bob Sloan
Alex Conconi	John Langford	Frank Stephan
Sanjoy Dasgupta	Christina Leslie	Christino Tamon
Tijl DeBie	Chi-jen Lin	Koji Tsuda
Nigel Duffy	Philip Long	Alexandre Tsybakov
Pierre Dupont	Fernando Lozano	Nicolas Vayatis
Ran El-Yaniv	Gabor Lugosi	Jean-Philippe Vert
Andre Elisseeff	Ulrike von Luxburg	Volodya Vovk
Eyal Even-Dar	Olvi Mangasarian	Grace Wahba
Theodoros Evgeniou	Shie Mannor	Osamu Watanabe
Amos Fiat	David McAllester	Rolf Wiehagen
Karim Filali	Ron Meir	Tong Zhang
Peter Flach	Shahar Mendelson	Sandra Zilles

Special thanks go to our local conference chairs: to Carl Smith, whose health problems would not let him finish the job (Carl, get better soon!), and to John Case, who thankfully agreed to take over this job. Together with the invaluable help of Cecilia Kullman, John saved us when things were gradually getting into a state of emergency. We would also like to thank Richard van de Stadt, Karin Bierig, and Sabrina Nielebock for help with the paper submissions and with the preparation of the proceedings, and Stephen Kwek for maintaining the conference Web site. Finally, we thank the National ICT Australia (www.nicta.com.au), the University of Maryland Institute for Advanced Computer Studies, and the Max Planck Institute for Biological Cybernetics for their generous sponsorship and administrative help.

June 2003

Bernhard Schölkopf (Max Planck Institute Tübingen)
Manfred K. Warmuth (University of California at Santa Cruz)
Program Chairs

Sponsors



Table of Contents

Target Area: Computational Game Theory

Tutorial: Learning Topics in Game-Theoretic Decision Making	1
<i>Michael L. Littman (Rutgers University)</i>	

Invited Talk

A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria	2
<i>Amy Greenwald and Amir Jafari (Brown University)</i>	

Contributed Talks

Preference Elicitation and Query Learning	13
<i>Avrim Blum (Carnegie Mellon University), Jeffrey C. Jackson (Duquesne University), Tuomas Sandholm, and Martin Zinkevich (Carnegie Mellon University)</i>	
Efficient Algorithms for Online Decision Problems	26
<i>Adam Kalai and Santosh Vempala (Massachusetts Institute of Technology)</i>	

Kernel Machines

Positive Definite Rational Kernels	41
<i>Corinna Cortes, Patrick Haffner, and Mehryar Mohri (AT&T Labs)</i>	
Bhattacharyya and Expected Likelihood Kernels	57
<i>Tony Jebara and Risi Kondor (Columbia University)</i>	
Maximal Margin Classification for Metric Spaces	72
<i>Matthias Hein and Olivier Bousquet (Max Planck Institute for Biological Cybernetics)</i>	
Maximum Margin Algorithms with Boolean Kernels	87
<i>Roni Khardon (Tufts University) and Rocco A. Servedio (Columbia University)</i>	
Knowledge-Based Nonlinear Kernel Classifiers	102
<i>Glenn M. Fung, Olvi L. Mangasarian, and Jude W. Shavlik (The University of Wisconsin-Madison)</i>	
Fast Kernels for Inexact String Matching	114
<i>Christina Leslie and Rui Kuang (Columbia University)</i>	

On Graph Kernels: Hardness Results and Efficient Alternatives	129
<i>Thomas Gärtner (Fraunhofer AIS, University of Bonn,</i> <i>and University of Bristol), Peter Flach (University of Bristol),</i> <i>and Stefan Wrobel (Fraunhofer AIS and University of Bonn)</i>	

Kernels and Regularization on Graphs	144
<i>Alexander J. Smola (The Australian National University)</i> <i>and Risi Kondor (Columbia University)</i>	

Data-Dependent Bounds for Multi-category Classification Based on Convex Losses	159
<i>Ilya Delyatnikov and Ron Meir (Technion)</i>	

Poster Session 1

Comparing Clusterings by the Variation of Information	173
<i>Marina Meilă (University of Washington)</i>	

Multiplicative Updates for Large Margin Classifiers	188
<i>Fei Sha, Lawrence K. Saul, and Daniel D. Lee</i> <i>(University of Pennsylvania)</i>	

Simplified PAC-Bayesian Margin Bounds	203
<i>David McAllester (Toyota Technological Institute at Chicago)</i>	

Sparse Kernel Partial Least Squares Regression	216
<i>Michinari Momma and Kristin P. Bennett</i> <i>(Rensselaer Polytechnic Institute)</i>	

Sparse Probability Regression by Label Partitioning	231
<i>Shantanu Chakrabartty, Gert Cauwenberghs</i> <i>(The John Hopkins University),</i> <i>and Jayadeva (Indian Institute of Technology)</i>	

Learning with Rigorous Support Vector Machines	243
<i>Jinbo Bi (Rensselaer Polytechnic Institute)</i> <i>and Vladimir N. Vapnik (NEC Laboratories America)</i>	

Robust Regression by Boosting the Median	258
<i>Balázs Kégl (University of Montreal)</i>	

Boosting with Diverse Base Classifiers	273
<i>Sanjoy Dasgupta (University of California at San Diego)</i> <i>and Philip M. Long (Genome Institute of Singapore)</i>	

Reducing Kernel Matrix Diagonal Dominance Using Semi-definite Programming	288
<i>Jaz Kandola (Royal Holloway University of London),</i> <i>Thore Graepel (Microsoft Research),</i> <i>and John Shawe-Taylor (Royal Holloway University of London)</i>	

Statistical Learning Theory

Optimal Rates of Aggregation	303
<i>Alexandre B. Tsybakov (Université Paris 6)</i>	
Distance-Based Classification with Lipschitz Functions	314
<i>Ulrike von Luxburg and Olivier Bousquet</i> <i>(Max Planck Institute for Biological Cybernetics)</i>	
Random Subclass Bounds	329
<i>Shahar Mendelson and Petra Philips</i> <i>(The Australian National University)</i>	
PAC-MDL Bounds	344
<i>Avrim Blum (Carnegie Mellon Univesity)</i> <i>and John Langford (IBM TJ Watson Research Center)</i>	

Online Learning

Universal Well-Calibrated Algorithm for On-Line Classification	358
<i>Vladimir Vovk (Royal Holloway University of London)</i>	
Learning Probabilistic Linear-Threshold Classifiers via Selective Sampling	373
<i>Nicolò Cesa-Bianchi, Alex Conconi (Università degli Studi di Milano),</i> <i>and Claudio Gentile (Università dell'Insubria)</i>	
Learning Algorithms for Enclosing Points in Bregmanian Spheres	388
<i>Koby Crammer and Yoram Singer</i> <i>(The Hebrew University of Jerusalem)</i>	
Internal Regret in On-Line Portfolio Selection	403
<i>Gilles Stoltz (Université Paris-Sud)</i> <i>and Gábor Lugosi (University Pompeu Fabra)</i>	

Other Approaches

Lower Bounds on the Sample Complexity of Exploration in the Multi-armed Bandit Problem	418
<i>Shie Mannor and John N. Tsitsiklis</i> <i>(Massachusetts Institute of Technology)</i>	

Smooth ϵ -Insensitive Regression by Loss Symmetrization	433
<i>Ofer Dekel, Shai Shalev-Shwartz, and Yoram Singer</i> (The Hebrew University of Jerusalem)	
On Finding Large Conjunctive Clusters	448
<i>Nina Mishra (HP Labs/Stanford University),</i> <i>Dana Ron (Tel-Aviv University), and Ram Swaminathan (HP Labs)</i>	
Learning Arithmetic Circuits via Partial Derivatives	463
<i>Adam R. Klivans (Harvard University) and Amir Shpilka</i> (Harvard University and Massachusetts Institute of Technology)	
Poster Session 2	
Using a Linear Fit to Determine Monotonicity Directions	477
<i>Malik Magdon-Ismail (RPI) and Joseph Sill (Plumtree Software)</i>	
Generalization Bounds for Voting Classifiers Based on Sparsity and Clustering	492
<i>Vladimir Koltchinskii (The University of New Mexico),</i> <i>Dmitry Panchenko (Massachusetts Institute of Technology),</i> <i>and Savina Andonova (Boston University)</i>	
Sequence Prediction Based on Monotone Complexity	506
<i>Marcus Hutter (IDSIA)</i>	
How Many Strings Are Easy to Predict?	522
<i>Yuri Kalnishkan, Vladimir Vovk, and Michael V. Vyugin</i> (Royal Holloway University of London)	
Polynomial Certificates for Propositional Classes	537
<i>Marta Arias, Roni Khardon (Tufts University),</i> <i>and Rocco A. Servedio (Columbia University)</i>	
On-Line Learning with Imperfect Monitoring	552
<i>Shie Mannor (Massachusetts Institute of Technology)</i> <i>and Nahum Shimkin (Technion)</i>	
Exploiting Task Relatedness for Multiple Task Learning	567
<i>Shai Ben-David (Corell University and Technion)</i> <i>and Reba Schuller (Cornell University)</i>	
Approximate Equivalence of Markov Decision Processes	581
<i>Eyal Even-Dar and Yishay Mansour (Tel-Aviv University)</i>	
An Information Theoretic Tradeoff between Complexity and Accuracy	595
<i>Ran Gilad-Bachrach, Amir Navot, and Naftali Tishby</i> (The Hebrew University of Jerusalem)	

Learning Random Log-Depth Decision Trees under the Uniform Distribution	610
<i>Jeffrey C. Jackson (Duchesne University)</i> <i>and Rocco A. Servedio (Columbia University)</i>	

Projective DNF Formulae and Their Revision	625
<i>Robert H. Sloan (University of Illinois at Chicago),</i> <i>Balázs Szörényi (University of Szeged),</i> <i>and György Turán (University of Illinois at Chicago)</i>	

Learning with Equivalence Constraints and the Relation to Multiclass Learning	640
<i>Aharon Bar-Hillel and Daphna Weinshall</i> <i>(The Hebrew University of Jerusalem)</i>	

Target Area: Natural Language Processing

Tutorial: Machine Learning Methods in Natural Language Processing	655
<i>Michael Collins (Massachusetts Institute of Technology)</i>	

Invited Talks

Learning from Uncertain Data	656
<i>Mehryar Mohri (AT&T Labs)</i>	
Learning and Parsing Stochastic Unification-Based Grammars	671
<i>Mark Johnson (Brown University)</i>	

Inductive Inference Learning

Generality's Price: Inescapable Deficiencies in Machine-Learned Programs	684
<i>John Case (University of Delaware),</i> <i>Keh-Jiann Chen (Academia Sinica),</i> <i>Sanjay Jain (National University of Singapore),</i> <i>Wolfgang Merkle (Universität Heidelberg),</i> <i>and James S. Royer (Syracuse University)</i>	

On Learning to Coordinate: Random Bits Help, Insightful Normal Forms, and Competency Isomorphisms	699
<i>John Case (University of Delaware),</i> <i>Sanjay Jain (National University of Singapore),</i> <i>Franco Montagna, Giulia Simi, and Andrea Sorbi (University of Siena)</i>	

Learning All Subfunctions of a Function	714
<i>Sanjay Jain (National University of Singapore),</i> <i>Efim Kinber (Sacred Heart University),</i> <i>and Rolf Wiehagen (University of Kaiserslautern)</i>	

Open Problems

When Is Small Beautiful? 729
Amiran Ambroladze (Lund University)
and John Shawe-Taylor (Royal Holloway University of London)

Learning a Function of r Relevant Variables 731
Avrim Blum (Carnegie Mellon University)

Subspace Detection: A Robust Statistics Formulation 734
Sanjoy Dasgupta (University of California at San Diego)

How Fast Is k -Means? 735
Sanjoy Dasgupta (University of California at San Diego)

Universal Coding of Zipf Distributions 736
Yoav Freund (Mitsubishi Electric Research Labs),
Alon Orlitsky, Prasad Santhanam, and Junan Zhang
(University of California at San Diego)

An Open Problem Regarding the Convergence
of Universal A Priori Probability 738
Marcus Hutter (IDSIA)

Entropy Bounds for Restricted Convex Hulls 741
Vladimir Koltchinskii (University of New Mexico)

Compressing to VC Dimension Many Points 743
Manfred K. Warmuth (University of California at Santa Cruz)

Author Index 745