

Contents

Preface / xiii

Nina Mishra and Tom Fawcett

ICML 2003 Organization / xv

ICML 2003 Corporate Sponsors / xviii

Hidden Markov Support Vector Machines / 3

Yasemin Altun, Ioannis Tsochantaridis, and Thomas Hofmann

Learning Distance Functions using Equivalence Relations / 11

Aharon Bar-Hillel, Tomer Hertz, Noam Shental, and Daphna Weinshall

Online Choice of Active Learning Algorithms / 19

Yoram Baram, Ran El-Yaniv, and Kobi Luz

Learning Logic Programs for Layout Analysis Correction / 27

Margherita Berardi, Michelangelo Ceci, Floriana Esposito, and Donato Malerba

Multi-Objective Programming in SVMs / 35

Jinbo Bi

Regression Error Characteristic Curves / 43

Jinbo Bi and Kristin P. Bennett

Choosing Between Two Learning Algorithms Based on Calibrated Tests / 51

Remco R. Bouckaert

Incorporating Diversity in Active Learning with Support Vector Machines / 59

Klaus Brinker

The Use of the Ambiguity Decomposition in Neural Network Ensemble Learning Methods / 67

Gavin Brown and Jeremy Wyatt

Tractable Bayesian Learning of Tree Augmented Naive Bayes Models / 75

Jesús Cerquides and Ramon López de Màntaras

AWESOME: A General Multiagent Learning Algorithm that Converges in Self-Play and
Learns a Best Response Against Stationary Opponents / 83

Vincent Conitzer and Tuomas Sandholm

BL-WoLF: A Framework For Loss-Bounded Learnability In Zero-Sum Games / 91

Vincent Conitzer and Tuomas Sandholm

Semi-Supervised Learning of Mixture Models / 99

Fabio Gagliardi Cozman, Ira Cohen, and Marcelo Cesar Cirelo

On Kernel Methods for Relational Learning / 107

Chad Cumby and Dan Roth

Fast Query-Optimized Kernel Machine Classification Via Incremental Approximate Nearest Support Vectors / 115
Dennis DeCoste and Dominic Mazzoni

Relational Instance Based Regression for Relational Reinforcement Learning / 123
Kurt Driessens and Jan Ramon

Design for an Optimal Probe / 131
Michael Duff

Diffusion Approximation for Bayesian Markov Chains / 139
Michael Duff

Using the Triangle Inequality to Accelerate k -Means / 147
Charles Elkan

Bayes Meets Bellman: The Gaussian Process Approach to Temporal Difference Learning / 154
Yaakov Engel, Shie Mannor, and Ron Meir

Action Elimination and Stopping Conditions for Reinforcement Learning / 162
Eyal Even-Dar, Shie Mannor, and Yishay Mansour

Utilizing Domain Knowledge in Neuroevolution / 170
James Fan, Raymond Lau, and Risto Miikkulainen

Boosting Lazy Decision Trees / 178
Xiaoli Zhang Fern and Carla E. Brodley

Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach / 186
Xiaoli Zhang Fern and Carla E. Brodley

The Geometry of ROC Space: Understanding Machine Learning Metrics through ROC Isometrics / 194
Peter A. Flach

An Analysis of Rule Evaluation Metrics / 202
Johannes Fürnkranz and Peter A. Flach

Margin Distribution and Learning / 210
Ashutosh Garg and Dan Roth

Perceptron Based Learning with Example Dependent and Noisy Costs / 218
Peter Geibel and Fritz Wysotzki

Hierarchical Policy Gradient Algorithms / 226
Mohammad Ghavamzadeh and Sridhar Mahadevan

Solving Noisy Linear Operator Equations by Gaussian Processes:
Application to Ordinary and Partial Differential Equations / 234
Thore Graepel

Correlated Q-Learning / 242
Amy Greenwald and Keith Hall

Online Ranking/Collaborative Filtering Using the Perceptron Algorithm / 250
Edward F. Harrington

Goal-directed Learning to Fly / 258
Andrew Isaac and Claude Sammut

Probabilistic Classifiers and the Concepts They Recognize / 266
Manfred Jaeger

Avoiding Bias when Aggregating Relational Data with Degree Disparity / 274	<i>David Jensen, Jennifer Neville, and Michael Hay</i>
A Faster Iterative Scaling Algorithm for Conditional Exponential Model / 282	<i>Rong Jin, Rong Yan, Jian Zhang, and Alex G. Hauptmann</i>
Transductive Learning via Spectral Graph Partitioning / 290	<i>Thorsten Joachims</i>
Evolving Strategies for Focused Web Crawling / 298	<i>Judy Johnson, Kostas Tsioutsoulis, and C. Lee Giles</i>
Exploration in Metric State Spaces / 306	<i>Sham Kakade, Michael Kearns, and John Langford</i>
Representational Issues in Meta-Learning / 313	<i>Alexandros Kalousis and Melanie Hilario</i>
Marginalized Kernels Between Labeled Graphs / 321	<i>Hisashi Kashima, Koji Tsuda, and Akihiro Inokuchi</i>
Informative Discriminant Analysis / 329	<i>Samuel Kaski and Jaakko Peltonen</i>
Characteristics of Long-term Learning in Soar and its Application to the Utility Problem / 337	<i>William G. Kennedy and Kenneth A. De Jong</i>
Unsupervised Learning with Permuted Data / 345	<i>Sergey Kirshner, Sridevi Parise, and Padhraic Smyth</i>
Discriminative Gaussian Mixture Models: A Comparison with Kernel Classifiers / 353	<i>Aldebaro Klautau, Nikola Jevtic, and Alon Orlitsky</i>
A Kernel Between Sets of Vectors / 361	<i>Risi Kondor and Tony Jebara</i>
The Significance of Temporal-Difference Learning in Self-Play Training TD-Rummy versus EVO-rummy / 369	<i>Clifford Kotnik and Jugul Kalita</i>
Visual Learning by Evolutionary Feature Synthesis / 376	<i>Krzysztof Krawiec and Bir Bhanu</i>
Classification of Text Documents Based on Minimum System Entropy / 384	<i>Raghu Krishnapuram, Krishna P. Chitrapura, and Sachindra Joshi</i>
Finding Underlying Connections: A Fast Graph-Based Method for Link Analysis and Collaboration Queries / 392	<i>Jeremy Kubica, Andrew Moore, David Cohn, and Jeff Schneider</i>
Learning with Idealized Kernels / 400	<i>James T. Kwok and Ivor W. Tsang</i>
The Pre-Image Problem in Kernel Methods / 408	<i>James Kwok and Ivor Tsang</i>
Improving Accuracy and Cost of Two-class and Multi-class Probabilistic Classifiers Using ROC Curves / 416	<i>Nicolas Lachiche and Peter Flach</i>
Reinforcement Learning as Classification: Leveraging Modern Classifiers / 424	<i>Michail Lagoudakis and Ronald Parr</i>

- Robust Induction of Process Models from Time-Series Data / 432
Pat Langley, Dileep George, Stephen Bay, and Kazumi Saito
- The Influence of Reward on the Speed of Reinforcement Learning: An Analysis of Shaping / 440
Adam Laud and Gerald DeJong
- Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression / 448
Wee Sun Lee and Bing Liu
- Linear Programming Boosting for Uneven Datasets / 456
Jurij Leskovec and John Shawe-Taylor
- Text Classification Using Stochastic Keyword Generation / 464
Cong Li, Ji-Rong Wen, and Hang Li
- A Loss Function Analysis for Classification Methods in Text Categorization / 472
Fan Li and Yiming Yang
- Decision Tree with Better Ranking / 480
Charles X. Ling and Robert J. Yan
- An Evaluation on Feature Selection for Text Clustering / 488
Tao Liu, Shengping Liu, Zheng Chen, and Wei-Ying Ma
- Link-based Classification / 496
Qing Lu and Lise Getoor
- Hierarchical Latent Knowledge Analysis for Co-occurrence Data / 504
Hiroshi Mamitsuka
- The Cross Entropy Method for Fast Policy Search / 512
Shie Mannor, Reuven Rubinstein, and Yohai Gat
- The Set Covering Machine with Data-Dependent Half-Spaces / 520
Mario Marchand, Mohak Shah, John Shawe-Taylor, and Marina Sokolova
- Identifying Predictive Structures in Relational Data Using Multiple Instance Learning / 528
Amy McGovern and David Jensen
- Planning in the Presence of Cost Functions Controlled by an Adversary / 536
H. Brendan McMahan, Geoffrey J. Gordon, and Avrim Blum
- Using Linear-threshold Algorithms to Combine Multi-class Sub-experts / 544
Chris Mesterharm
- Optimal Reinsertion: A New Search Operator for Accelerated and More Accurate Bayesian Network Structure Learning / 552
Andrew Moore and Weng-Keen Wong
- Error Bounds for Approximate Policy Iteration / 560
Rémi Munos
- Machine Learning with Hyperkernels / 568
Cheng Soon Ong and Alexander J. Smola
- Justification-based Multiagent Learning / 576
Santi Ontañón and Enric Plaza
- Mixtures of Conditional Maximum Entropy Models / 584
Dmitry Pavlov, Alexandrin Popescul, David M. Pennock, and Lyle H. Ungar

- Online Feature Selection using Grafting / 592
Simon Perkins and James Theiler
- Weighted Order Statistic Classifiers with Large Rank-Order Margin / 600
Reid Porter, Damian Eads, Don Hush, and James Theiler
- Relativized Options: Choosing the Right Transformation / 608
Balaraman Ravindran and Andrew G. Barto
- Tackling the Poor Assumptions of Naive Bayes Text Classifiers / 616
Jason D. Rennie, Lawrence Shih, Jaime Teevan, and David Karger
- Learning with Knowledge from Multiple Experts / 624
Matthew Richardson and Pedro Domingos
- Combining TD-learning with Cascade-correlation Networks / 632
François Rivest and Doina Precup
- Kernel PLS-SVC for Linear and Nonlinear Classification / 640
Roman Rosipal, Leonard J. Trejo, and Bryan Matthews
- Stochastic Local Search in k -Term DNF Learning / 648
Ulrich Rückert and Stefan Kramer
- Q-Decomposition for Reinforcement Learning Agents / 656
Stuart Russell and Andrew L. Zimdars
- Adaptive Overrelaxed Bound Optimization Methods / 664
Ruslan Salakhutdinov and Sam Roweis
- Optimization with EM and Expectation-Conjugate-Gradient / 672
Ruslan Salakhutdinov, Sam Roweis, and Zoubin Ghahramani
- TD(0) Converges Provably Faster than the Residual Gradient Algorithm / 680
Ralf Schoknecht and Artur Merke
- On State Merging in Grammatical Inference: A Statistical Approach for Dealing with Noisy Data / 688
Marc Sebban and Jean-Christophe Janodet
- Text Bundling: Statistics Based Data-Reduction / 696
Lawrence Shih, Jason Rennie, Yu-Han Chang, and David Karger
- Flexible Mixture Model for Collaborative Filtering / 704
Luo Si and Rong Jin
- Learning Predictive State Representations / 712
Satinder Singh, Michael L. Littman, Nicholas K. Jong, David Pardoe, and Peter Stone
- Weighted Low-Rank Approximations / 720
Nathan Srebro and Tommi Jaakkola
- Learning To Cooperate in a Social Dilemma: A Satisficing Approach to Bargaining / 728
Jeffrey L. Stimpson and Michael A. Goodrich
- Evolutionary MCMC Sampling and Optimization in Discrete Spaces / 736
Malcolm J. A. Strens
- Learning on the Test Data: Leveraging Unseen Features / 744
Ben Taskar, Ming Fai Wong, and Daphne Koller

Low Bias Bagged Support Vector Machines / 752

Giorgio Valentini and Thomas G. Dietterich

SimpleSVM / 760

S. V. N. Vishwanathan, Alexander J. Smola, and M. Narashima Murty

Testing Exchangeability On-Line / 768

Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman

Model-based Policy Gradient Reinforcement Learning / 776

Xin Wang and Thomas G. Dietterich

Learning Mixture Models with the Latent Maximum Entropy Principle / 784

Shaojun Wang, Dale Schuurmans, Fuchun Peng, and Yunxin Zhao

Principled Methods for Advising Reinforcement Learning Agents / 792

Eric Wiewiora, Garrison Cottrell, and Charles Elkan

DISTILL: Learning Domain-Specific Planners by Example / 800

Elly Winner and Manuela Veloso

Bayesian Network Anomaly Pattern Detection for Disease Outbreaks / 808

Weng-Keen Wong, Andrew Moore, Gregory Cooper, and Michael Wagner

Adaptive Feature-Space Conformal Transformation for Imbalanced-Data Learning / 816

Gang Wu and Edward Y. Chang

New ν -Support Vector Machines and their Sequential Minimal Optimization / 824

Xiaoyun Wu and Rohini Srihari

Cross-Entropy Directed Embedding of Network Data / 832

Takeshi Yamada, Kazumi Saito, and Naonori Ueda

Decision-tree Induction from Time-series Data Based on a Standard-example Split Test / 840

Yuu Yamada, Einoshin Suzuki, Hideto Yokoi, and Katsuhiko Takabayashi

Optimizing Classifier Performance via an Approximation to the Wilcoxon-Mann-Whitney Statistic / 848

Lian Yan, Robert Dodier, Michael C. Mozer, and Richard Wolniewicz

Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution / 856

Lei Yu and Huan Liu

Isometric Embedding and Continuum ISOMAP / 864

Hongyuan Zha, and Zhenyue Zhang

Learning Metrics via Discriminant Kernels and Multidimensional Scaling:

Toward Expected Euclidean Representation / 872

Zhihua Zhang

Learning Decision Tree Classifiers from Attribute Value Taxonomies and Partially Specified Data / 880

Jun Zhang and Vasant Honavar

Modified Logistic Regression: An Approximation to SVM and Its

Applications in Large-Scale Text Categorization / 888

Jian Zhang, Rong Jin, Yiming Yang, and Alex G. Hauptmann

Exploration and Exploitation in Adaptive Filtering Based on Bayesian Active Learning / 896

Yi Zhang, Wei Xu, and Jamie Callan

On the Convergence of Boosting Procedures / 904

Tong Zhang and Bin Yu

Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions / 912

Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty

Eliminating Class Noise in Large Datasets / 920

Xingquan Zhu, Xindong Wu, and Qijun Chen

Online Convex Programming and Generalized Infinitesimal Gradient Ascent / 928

Martin Zinkevich

Index / 937