Contents

Preface
Acknowledgements
Knowledge Representation Issues in Control Knowledge Learning
   Ricardo Aler, Daniel Borrajo, and Pedro Isasi .............................................. 1
Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers
   Erin L. Allwein, Robert E. Schapire, and Yoram Singer .................................... 9
A Nonparametric Approach to Noisy and Costly Optimization
   Brigham S. Anderson, Andrew W. Moore, and David Cohn .................................. 17
Behavioral Cloning of Student Pilots with Modular Neural Networks
   Charles W. Anderson, Bruce A. Draper, and David A. Peterson ............................ 25
Combining Multiple Perspectives
   Bikramit Banerjee, Sandip Debnath, and Sandip Sen ........................................ 33
Reinforcement Learning in POMDP’s via Direct Gradient Ascent
   Jonathan Baxter and Peter L. Bartlett ............................................................. 41
Characterizing Model Errors and Differences
   Stephen D. Bay and Michael J. Pazzani ......................................................... 49
Duality and Geometry in SVM Classifiers
   Kristin P. Bennett and Erin J. Bredensteiner ................................................... 57
A Column Generation Algorithm For Boosting
   Kristin P. Bennett, Ayhan Demiriz, and John Shawe-Taylor ................................. 65
Disciple-COA: From Agent Programming to Agent Teaching
   M. Boicu, G. Tecuci, D. Marcu, M. Bowman, P. Shyr, F. Ciucu, and C. Levcovici .......... 73
Classification of Individuals with Complex Structure
   A. F. Bowers, C. Giraud-Carrier, and J. W. Lloyd ............................................ 81
Convergence Problems of General-Sum Multiagent Reinforcement Learning
   Michael Bowling .................................................................................................. 89
Finding Variational Structure in Data by Cross-Entropy Optimization
   Matthew Brand ................................................................................................. 95
Challenges of the Email Domain for Text Classification
   Jake D. Brutlag and Christopher Meek ............................................................. 103
Query Learning with Large Margin Classifiers
   Colin Campbell, Nello Cristianini, and Alex Smola ............................................. 111
Dimension Reduction Techniques for Training Polynomial Networks
   William M. Campbell, Kari Torkkola, and Sreeram V. Balakrishnan ......................... 119
Learning to Create Customized Authority Lists
Huan Chang, David Cohn, and Andrew K. McCallum ......................................................... 127

Learning to Select Text Databases with Neural Nets
Yong S. Choi and Suk I. Yoo ....................................................................................... 135

A Divide and Conquer Approach to Learning from Prior Knowledge
Eric Chown and Thomas G. Dietterich ........................................................................ 143

Learning in Non-stationary Conditions: A Control Theoretic Approach
Jefferson Coelho and Rod Grupen ............................................................................... 151

Automatically Extracting Features for Concept Learning from the Web
William W. Cohen ...................................................................................................... 159

Learning to Probabilistically Identify Authoritative Documents
David Cohn and Huan Chang ..................................................................................... 167

Discriminative Reranking for Natural Language Parsing
Michael Collins ........................................................................................................... 175

Automatic Identification of Mathematical Concepts
Simon Colton, Alan Bundy, and Toby Walsh ................................................................ 183

On-line Learning for Humanoid Robot Systems
Jörg Conradt, Gaurav Tevatia, Sethu Vijayakumar, and Stefan Schaal ..................... 191

Using Multiple Levels of Learning and Diverse Evidence to Uncover Coordinately Controlled Genes
Mark Craven, David Page, Jude Shavlik, Joseph Bockhorst, and Jeremy Glasner ....... 199

Fixed Points of Approximate Value Iteration and Temporal-Difference Learning
Daniela Pucci de Farias and Benjamin Van Roy ......................................................... 207

Hidden Strengths and Limitations: An Empirical Investigation of Reinforcement Learning
Gerald DeJong .............................................................................................................. 215

Bayesian Averaging of Classifiers and the Overfitting Problem
Pedro Domingos ......................................................................................................... 223

A Unified Bias-Variance Decomposition and its Applications
Pedro Domingos ......................................................................................................... 231

Exploiting the Cost (In)sensitivity of Decision Tree Splitting Criteria
Chris Drummond and Robert C. Holte ........................................................................ 239

Feature Subset Selection and Order Identification for Unsupervised Learning
Jennifer G. Dy and Carla E. Brodley ........................................................................... 247

Anomaly Detection over Noisy Data using Learned Probability Distributions
Eleazar Eskin ............................................................................................................... 255

Ideal Theory Refinement under Object Identity
Floriana Esposito, Nicola Fanizzi, Stefano Ferilli, and Giovanni Semeraro ............... 263

Bounds on the Generalization Performance of Kernel Machine Ensembles
Theodoros Evgeniou, Luis Perez-Breva, Massimiliano Pontil, and Tomaso Poggio .... 271

Online Ensemble Learning: An Empirical Study
Alan Fern and Robert Givan ..................................................................................... 279

Learning Subjective Functions with Large Margins
Claude-Nicolas Fiechter and Seth Rogers .................................................................... 287
Relative Loss Bounds for Temporal-Difference Learning
Jurgen Forster and Manfred Warmuth .......................................................... 295

Using Error-Correcting Codes for Text Classification
Rayid Ghani .................................................................................................. 303

Analyzing Relational Learning in the Phase Transition Framework
Attilio Giordana, Lorenzo Saitta, Michele Sebag, and Marco Botta .................. 311

Learning Multiple Models for Reward Maximization
Dani Goldberg and Maja J. Matarić ............................................................... 319

Enhancing Supervised Learning with Unlabeled Data
Sally Goldman and Yan Zhou ......................................................................... 327

Learning Filaments
Geoffrey J. Gordon and Andrew Moore ....................................................... 335

Localizing Policy Gradient Estimates to Action Transitions
Gregory Z. Grudic and Lyle H. Ungar .......................................................... 343

Learning Curved Multinomial Subfamilies for Natural Language Processing and Information Retrieval
Keith Hall and Thomas Hofmann ................................................................... 351

Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning
Mark A. Hall ................................................................................................. 359

Empirical Bayes for Learning to Learn
Tom Heskes .................................................................................................. 367

Meta-Learning for Phonemic Annotation of Corpora
Veronique Hoste, Walter Daelemans, Erik Tjong Kim Sang, and Steven Gillis .. 375

An Integrated Connectionist Approach to Reinforcement Learning for Robotic Control
Dean F. Hougen, Maria Gini, and James Slagle ............................................... 383

Data as Ensembles of Records: Representation and Comparison
Nicholas R. Howe ......................................................................................... 391

Why Discretization Works for Naive Bayesian Classifiers
Chun-Nan Hsu, Hung-Ju Huang, and Tzu-Tsung Wong .................................. 399

Experimental Results on Q-Learning for General-Sum Stochastic Games
Junling Hu and Michael P. Wellman ............................................................ 407

Learning Declarative Control Rules for Constraint-Based Planning
Yi-Cheng Huang, Bart Selman, and Henry Kautz ........................................... 415

Approximate Dimension Equalization in Vector-based Information Retrieval
Fan Jiang and Michael L. Littman .................................................................... 423

Estimating the Generalization Performance of an SVM Efficiently
Thorsten Joachims ....................................................................................... 431

State-based Classification of Finger Gestures from Electromyographic Signals
Peter Ju, Leslie Pack Kaelbling, and Yoram Singer ....................................... 439

A Universal Generalization for Temporal-Difference Learning Using Haar Basis Functions
Susumu Katayama, Hajime Kimura, and Shigenobu Kobayashi ....................... 447

MultiStage Cascading of Multiple Classifiers: One Man's Noise is Another Man's Data
Cenk Kaynak and Ethem Alpaydın ............................................................... 455
Pseudo-convergent Q-Learning by Competitive Pricebots

Jeffrey O. Kephart and Gerald J. Tesauro ........................................... 463

Learning Horn Expressions with LogAn-H

Roni Khardon ......................................................................................... 471

Learning Bayesian Networks for Diverse and Varying Numbers of Evidence Sets

ZuWhan Kim and Ramakant Nevatia ....................................................... 479

Detecting Concept Drift with Support Vector Machines

Ralf Klinkenberg and Thorsten Joachims .............................................. 487

A Dynamic Adaptation of AD-trees for Efficient Machine Learning on Large Data Sets

Paul Komarek and Andrew Moore ......................................................... 495

Voting Nearest-Neighbor Subclassifiers

Miroslav Kubat and Martin Cooperon, Jr. ............................................. 503

Algorithm Selection using Reinforcement Learning

Michail G. Lagoudakis and Michael L. Littman .................................... 511

Data Reduction Techniques for Instance-Based Learning from Human/Computer Interface Data

Terran Lane and Carla E. Brodley ......................................................... 519

Version Space Algebra and its Application to Programming by Demonstration

Tessa Lau, Pedro Domingos, and Daniel S. Weld ................................... 527

An Algorithm for Distributed Reinforcement Learning in Cooperative Multi-Agent Systems

Martin Lauer and Martin Riedmiller ...................................................... 535

A Bayesian Approach to Temporal Data Clustering using Hidden Markov Models

Cen Li and Gautam Biswas ................................................................. 543

The Space of Jumping Emerging Patterns and Its Incremental Maintenance Algorithms

Jinyan Li, Kotagiri Ramamohanarao, and Guozhu Dong ......................... 551

Selective Voting for Perceptron-like Online Learning

Yi Li ........................................................................................................ 559

An Initial Study of an Adaptive Hierarchical Vision System

Marcus A. Maloof .................................................................................. 567

Efficient Mining from Large Databases by Query Learning

Hiroshi Mamitsuka and Naoki Abe ...................................................... 575

Bootstrap Methods for the Cost-Sensitive Evaluation of Classifiers

Dragos D. Margineantu and Thomas G. Dietterich ................................ 583

Maximum Entropy Markov Models for Information Extraction and Segmentation

Andrew McCallum, Dayne Freitag, and Fernando Pereira ..................... 591

Mixtures of Factor Analyzers

Geoffrey McLachlan and David Peel .................................................... 595

“Boosting” a Positive-Data-Only Learner

Andrew Mitchell .................................................................................... 607

Machine Learning for Subproblem Selection

Robert Moll, Theodore J. Perkins, and Andrew G. Barto ....................... 615

Acquisition of Stand-up Behavior by a Real Robot using Hierarchical Reinforcement Learning

Jun Morimoto and Kenji Doya ............................................................ 623
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Chomsky-like Grammars for Biological Sequence Families</td>
<td>631</td>
</tr>
<tr>
<td>S. H. Muggleton, C. H. Bryant, and A. Srinivasan</td>
<td></td>
</tr>
<tr>
<td>Complete Cross-Validation for Nearest Neighbor Classifiers</td>
<td>639</td>
</tr>
<tr>
<td>Matthew Mulin and Rahul Sukthankar</td>
<td></td>
</tr>
<tr>
<td>Rates of Convergence for Variable Resolution Schemes in Optimal Control</td>
<td>647</td>
</tr>
<tr>
<td>Rémi Munos and Andrew W. Moore</td>
<td></td>
</tr>
<tr>
<td>A Boosting Approach to Topic Spotting on Subdialogues</td>
<td>655</td>
</tr>
<tr>
<td>Kary Myers, Michael Kearns, Satinder Singh, and Marilyn A. Walker</td>
<td></td>
</tr>
<tr>
<td>Algorithms for Inverse Reinforcement Learning</td>
<td>663</td>
</tr>
<tr>
<td>Andrew Y. Ng and Stuart Russell</td>
<td></td>
</tr>
<tr>
<td>Learning Probabilistic Models for Decision-Theoretic Navigation of Mobile Robots</td>
<td>671</td>
</tr>
<tr>
<td>Daniel Nikovski and Illah Nourbakhsh</td>
<td></td>
</tr>
<tr>
<td>An Approach to Data Reduction and Clustering with Theoretical Guarantees</td>
<td>679</td>
</tr>
<tr>
<td>Partha Niyogi and Narendra Karmarkar</td>
<td></td>
</tr>
<tr>
<td>Comparing the Minimum Description Length Principle and Boosting in the Analysis of Discourse</td>
<td>687</td>
</tr>
<tr>
<td>Tadashi Nomoto and Yuji Matsumoto</td>
<td></td>
</tr>
<tr>
<td>Generalized Average-Case Analyses of the Nearest Neighbor Algorithm</td>
<td>695</td>
</tr>
<tr>
<td>Seishi Okamoto and Nobuhiro Yagami</td>
<td></td>
</tr>
<tr>
<td>FeatureBoost: A Meta-Learning Algorithm that Improves Model Robustness</td>
<td>703</td>
</tr>
<tr>
<td>Joseph O'Sullivan, John Langford, Rich Caruana, and Avrim Blum</td>
<td></td>
</tr>
<tr>
<td>Learning Distributed Representations by Mapping Concepts and Relations into a Linear Space</td>
<td>711</td>
</tr>
<tr>
<td>Alberto Paccanaro and Geoffrey E. Hinton</td>
<td></td>
</tr>
<tr>
<td>Clustering the Users of Large Web Sites into Communities</td>
<td>719</td>
</tr>
<tr>
<td>Georgios Paliouras, Christos Papatheodorou, Vangelis Karkaletsis, and Constantine Spyropoulos</td>
<td></td>
</tr>
<tr>
<td>X-means: Extending K-means with Efficient Estimation of the Number of Clusters</td>
<td>727</td>
</tr>
<tr>
<td>Dan Pelleg and Andrew Moore</td>
<td></td>
</tr>
<tr>
<td>A Normative Examination of Ensemble Learning Algorithms</td>
<td>735</td>
</tr>
<tr>
<td>David M. Pennock, Pedrito Maynard-Reid II, C. Lee Giles, and Eric Horvitz</td>
<td></td>
</tr>
<tr>
<td>Meta-Learning by Landmarking Various Learning Algorithms</td>
<td>743</td>
</tr>
<tr>
<td>Bernhard Pfahringer, Hilan Bensusan, and Christophe Giraud-Carrier</td>
<td></td>
</tr>
<tr>
<td>Constructive Feature Learning and the Development of Visual Expertise</td>
<td>751</td>
</tr>
<tr>
<td>Justus H. Piater and Roderic A. Grupen</td>
<td></td>
</tr>
<tr>
<td>Eligibility Traces for Off-Policy Policy Evaluation</td>
<td>759</td>
</tr>
<tr>
<td>Doina Precup, Richard S. Sutton, and Satinder Singh</td>
<td></td>
</tr>
<tr>
<td>Shaping in Reinforcement Learning by Changing the Physics of the Problem</td>
<td>767</td>
</tr>
<tr>
<td>Jette Randlov</td>
<td></td>
</tr>
<tr>
<td>Combining Reinforcement Learning with a Local Control Algorithm</td>
<td>775</td>
</tr>
<tr>
<td>Jette Randlov, Andrew G. Barto, and Michael T. Rosenstein</td>
<td></td>
</tr>
<tr>
<td>Adaptive Resolution Model-Free Reinforcement Learning: Decision Boundary Partitioning</td>
<td>783</td>
</tr>
<tr>
<td>Stuart I. Reynolds</td>
<td></td>
</tr>
<tr>
<td>Knowledge Propagation in Model-based Reinforcement Learning Tasks</td>
<td>791</td>
</tr>
<tr>
<td>Corinna Richter and Jörg Stachowiak</td>
<td></td>
</tr>
</tbody>
</table>
Image Color Constancy Using EM and Cached Statistics
Charles Rosenberg .................................................. 799

Learning to Fly: An Application of Hierarchical Reinforcement Learning
Malcolm Ryan and Mark Reid ....................................... 807

Direct Bayes Point Machines
Matthias Rychetsky, John Shawe-Taylor, and Manfred Glesner ................. 815

Achieving Efficient and Cognitively Plausible Learning in Backgammon
Scott Sanner, John R. Anderson, Christian Lebiere, and Marsha Lovett .......... 823

Predicting the Generalization Performance of Cross Validatory Model Selection Criteria
Tobias Scheffer ...................................................... 831

Less is More: Active Learning with Support Vector Machines
Greg Schohn and David Cohn ...................................... 839

An Adaptive Regularization Criterion for Supervised Learning
Dale Schuurmans and Finnegan Southey ............................ 847

Instance Pruning as an Information Preserving Problem
Marc Sebban and Richard Nock ..................................... 855

Incremental Learning in SwiftFile
Richard B. Segal and Jeffrey O. Kephart ................................ 863

Using Knowledge to Speed Learning: Knowledge-based Cascade-correlation and Multi-task Learning
Thomas R. Shultz and Francois Rivest ................................ 871

Obtaining Simplified Rule Bases by Hybrid Learning
Ricardo Bezerra de Andrade e Silva and Teresa Bernarda Ludermir ............... 879

Learning to Predict Performance from Formula Modeling and Training Data
Bryan Singer and Manuela Veloso .................................. 887

Discovering Test Set Regularities in Relational Domains
Sean Slattery and Tom Mitchell ..................................... 895

Practical Reinforcement Learning in Continuous Spaces
William D. Smart and Leslie Pack Kaelbling .......................... 903

Sparse Greedy Matrix Approximation for Machine Learning
Alex J. Smola and Bernhard Schölkopf ................................ 911

Using Learning by Discovery to Segment Remotely Sensed Images
Leen-Kiat Soh and Costas Tsatsoulis ................................ 919

Multi-agent Q-learning and Regression Trees for Automated Pricing Decisions
Manu Sridharan and Gerald Tesuaro .................................. 927

TPOT-RL Applied to Network Routing
Peter Stone ......................................................... 935

A Bayesian Framework for Reinforcement Learning
Malcolm Strens ....................................................... 943

Feature Selection and Incremental Learning of Probabilistic Concept Hierarchies
Luis Talavera ......................................................... 951

Efficient Learning Through Evolution: Neural Programming and Internal Reinforcement
Astro Teller and Manuela Veloso .................................... 959
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection of Support Vector Kernel Parameters for Improved Generalization</td>
<td>Loo-Nin Teow and Kia-Fock Loe</td>
<td>967</td>
</tr>
<tr>
<td>Probabilistic DFA Inference using Kullback-Leibler Divergence and Minimality</td>
<td>Franck Thollard, Pierre Dupont, and Colin de la Higuera</td>
<td>975</td>
</tr>
<tr>
<td>A Comparative Study of Cost-Sensitive Boosting Algorithms</td>
<td>Kai Ming Ting</td>
<td>983</td>
</tr>
<tr>
<td>Discovering the Structure of Partial Differential Equations from Example Behaviour</td>
<td>Ljupco Todorovski, Saso Dzeroski, Ashwin Srinivasan, Jonathan Whiteley, and David Gavaghan</td>
<td>991</td>
</tr>
<tr>
<td>Support Vector Machine Active Learning with Applications to Text Classification</td>
<td>Simon Tong and Daphne Koller</td>
<td>999</td>
</tr>
<tr>
<td>Partial Linear Trees</td>
<td>Luis Torgo</td>
<td>1007</td>
</tr>
<tr>
<td>Mutual Information in Learning Feature Transformations</td>
<td>Kari Yrkkola and William M. Campbell</td>
<td>1015</td>
</tr>
<tr>
<td>Local Expert Autoassociators for Anomaly Detection</td>
<td>Geoffrey G. Towell</td>
<td>1023</td>
</tr>
<tr>
<td>Learning Priorities From Noisy Examples</td>
<td>Geoffrey G. Towell, Thomas Petsche, and Michael R. Miller</td>
<td>1031</td>
</tr>
<tr>
<td>Hierarchical Unsupervised Learning</td>
<td>Shivakumar Vaithyanathan and Byron Dom</td>
<td>1039</td>
</tr>
<tr>
<td>Model Selection Criteria for Learning Belief Nets: An Empirical Comparison</td>
<td>Tim Van Allen and Russ Greiner</td>
<td>1047</td>
</tr>
<tr>
<td>Unpacking Multi-valued Symbolic Features and Classes in Memory-Based Language Learning</td>
<td>Antal van den Bosch and Jakub Zavrel</td>
<td>1055</td>
</tr>
<tr>
<td>Bootstrapping Syntax and Recursion using Alignment-Based Learning</td>
<td>Menno van Zaanen</td>
<td>1063</td>
</tr>
<tr>
<td>An Evolutionary Approach to Evidence-Based Learning of Deterministic Finite Automata</td>
<td>Stefan Veeser</td>
<td>1071</td>
</tr>
<tr>
<td>Locally Weighted Projection Regression: Incremental Real Time Learning in High Dimensional Space</td>
<td>Sethu Vijayakumar and Stefan Schaal</td>
<td>1079</td>
</tr>
<tr>
<td>A Quantification of Distance Bias Between Evaluation Metrics In Classification</td>
<td>Ricardo Vilalta and Daniel Oblinger</td>
<td>1087</td>
</tr>
<tr>
<td>Discovering Homogeneous Regions in Spatial Data through Competition</td>
<td>Slobodan Vucetic and Zoran Obradovic</td>
<td>1095</td>
</tr>
<tr>
<td>Clustering with Instance-level Constraints</td>
<td>Kiri Wagstaff and Claire Cardie</td>
<td>1103</td>
</tr>
<tr>
<td>Using Natural Language Processing and Discourse Features to Identify Understanding Errors</td>
<td>Marilyn Walker, Jerry Wright, and Irene Langkilde</td>
<td>1111</td>
</tr>
<tr>
<td>Solving the Multiple-Instance Problem: A Lazy Learning Approach</td>
<td>Jun Wang and Jean-Daniel Zucker</td>
<td>1119</td>
</tr>
<tr>
<td>Enhancing the Plausibility of Law Equation Discovery</td>
<td>Takashi Washio, Hiroshi Motoda, and Yuji Niwa</td>
<td>1127</td>
</tr>
</tbody>
</table>
Lightweight Rule Induction
   Sholom M. Weiss and Nitin Indurkhya ................................................................. 1135

Classification with Multiple Latent Variable Models using Maximum Entropy Discrimination
   Machiel Westerdijk and Wim Wiegerinck ............................................................. 1143

Multi-Agent Reinforcement Learning for Traffic Light Control
   Marco Wiering ......................................................................................................... 1151

The Effect of the Input Density Distribution on Kernel-based Classifiers
   Christopher K. I. Williams and Matthias Seeger ..................................................... 1159

Combining Multiple Learning Strategies for Effective Cross Validation
   Yiming Yang, Thomas Ault, and Thomas Pierce ..................................................... 1167

Linear Discriminant Trees
   Olcay Taner Yildiz and Ethem Alpaydin ............................................................... 1175

Improving Short-Text Classification using Unlabeled Background Knowledge
   Sarah Zelikovitz and Haym Hirsh .......................................................................... 1183

A Probability Analysis on the Value of Unlabeled Data for Classification Problems
   Tong Zhang and Frank J. Oles ................................................................................. 1191

Induction of Concept Hierarchies from Noisy Data
   Blaž Zupan, Ivan Bratko, Marko Bohanec, and Janez Demšar ............................... 1199

Crafting Papers on Machine Learning
   Pat Langley ............................................................................................................ 1207

Author Index ............................................................................................................. 1217