

Contents

An Ensemble Pruning Primer	1
<i>Grigorios Tsoumakas, Ioannis Partalas, Ioannis Vlahavas</i>	
1 Introduction	1
2 Background	2
2.1 Producing the Models	2
2.2 Combining the Models	3
3 A Taxonomy of Ensemble Pruning Methods	4
4 Ranking-Based Methods	4
5 Clustering-Based Methods	5
6 Optimization-Based Methods	6
6.1 Genetic Algorithms	6
6.2 Semi-definite Programming	6
6.3 Hill Climbing	7
7 Other Methods	10
7.1 Statistical Procedures	10
7.2 Reinforcement Learning	10
7.3 Boosting	11
8 Conclusions	11
References	12
 Evade Hard Multiple Classifier Systems	15
<i>Battista Biggio, Giorgio Fumera, Fabio Roli</i>	
1 Introduction	15
2 Related Work	17
2.1 Previous Works on Multiple Classifiers for Security Applications	17
2.2 A Theoretical Framework for Adversarial Classification Problems	19
3 Are Multiple Classifier Systems Harder to Evade?	21

3.1	Adding Features to a Classification System	22
3.2	Splitting Features across an Ensemble of Classifiers	24
4	A Case Study in Spam Filtering	28
4.1	Adding Features to a Spam Filter	29
4.2	Splitting the Features of a Spam Filter across an Ensemble of Classifiers	32
5	Conclusions	37
	References	37
 A Personal Antispam System Based on a Behaviour-Knowledge Space Approach		39
<i>Francesco Gargiulo, Antonio Penta, Antonio Picariello, Carlo Sansone</i>		
1	Introduction	39
2	Related Work	40
3	System Architecture	42
4	Textual Features	43
4.1	Semantic Features	43
4.2	Syntactic Features	46
5	Image Features	47
5.1	Visual Features	47
5.2	OCR-Based Features	49
6	Combining Text-Based and Image-Based Classifiers	50
7	Experimental Results	52
8	Conclusion	55
	References	56
 Weighted Decoding ECOC for Facial Action Unit Classification		59
<i>Terry Windeatt</i>		
1	Introduction	59
2	Ensembles and Bootstrapping	61
3	Error-Correcting Output Coding ECOC	63
3.1	Motivation	64
3.2	ECOC Algorithm and OOB Estimate	65
3.3	Coding Strategies and Errors	66
3.4	Weighted Decoding	68
4	Dataset and Feature Extraction	69
5	Experiments on Cohn-Kanade Database	71
6	Discussion	74
7	Conclusion	75
	References	75

Prediction of Gene Function Using Ensembles of SVMs and Heterogeneous Data Sources	79
<i>Matteo Re, Giorgio Valentini</i>	
1 Introduction	79
2 Methods	80
2.1 Linear Weighted Combination with Linear and Logarithmic Weights	81
2.2 Decision Templates	82
3 Experimental Setup	83
3.1 Heterogeneous Biomolecular Datasets	83
3.2 The Functional Catalogue (FunCat)	85
3.3 Base Learners Tuning and Generation of Optimized Classifiers	86
4 Results	86
5 Discussion	88
6 Conclusions	89
References	90
Partitioner Trees for Classification: A New Ensemble Method	93
<i>Georg Kreml, Vera Hofer</i>	
1 Introduction	93
2 Partitioner Trees	95
2.1 Algorithm	95
2.2 Discussion	99
3 Experiments	104
3.1 Experimental Setup	104
3.2 Results	105
3.3 Conclusion	111
References	112
Disturbing Neighbors Diversity for Decision Forests	113
<i>Jesús Maudes, Juan J. Rodríguez, César García-Osorio</i>	
1 Introduction	113
2 Method	115
3 Results	119
4 Lesion Study	124
5 Conclusion	131
References	132
Improving Supervised Learning with Multiple Clusterings ...	135
<i>Cédric Wemmert, Germain Forestier, Sébastien Dervaux</i>	
1 Introduction	135
2 Related Works	136
3 Description of the Method	138

3.1	Improving Supervised Classification with Clustering	138
3.2	The Proposed Method	140
4	Experiments	142
4.1	Artificial Benchmark Evaluation	142
4.2	Real Data Evaluation	144
5	Conclusion	148
	References	148
The Neighbors Voting Algorithm and Its Applications		151
<i>Gabriele Lombardi, and Elena Casiraghi, Paola Campadelli</i>		
1	Introduction	151
2	The Tensor Voting Framework	152
3	The Neighbors Voting Algorithm	155
3.1	Statistical Interpretation	156
4	Applications of the Neighbors Voting Algorithm	157
4.1	Point Clustering	157
4.2	Classification	158
4.3	Image Inpainting with NV	160
5	Results	161
5.1	Clustering Results	161
5.2	Classification Results	165
5.3	Inpainting Results	170
6	Conclusions and Future Work	170
	References	172
Clustering Ensembles with Active Constraints		175
<i>Muna Al-Razgan, Carlotta Domeniconi</i>		
1	Introduction	175
2	Related Work	176
3	Locally Adaptive Clustering	177
4	Selecting Informative Constraints	179
5	Chunklet Graph	181
6	Chunklet Assignment	182
7	Constrained-Weighted Bipartite Partitioning Algorithm (C-WBPA)	182
8	Empirical Evaluation	185
8.1	Analysis of the Results	188
9	Conclusions	188
	References	189
Verifiable Ensembles of Low-Dimensional Submodels for Multi-class Problems with Imbalanced Misclassification Costs		191
<i>Sebastian Nusser, Clemens Otte, Werner Hauptmann</i>		
1	Introduction	191

2	The Binary Ensemble Framework	193
2.1	Decision Tree-Like Ensemble Model	194
2.2	Non-Hierarchical Ensemble Model	194
2.3	An Illustrative Example	195
3	Multi-class Extensions of Binary Classifiers	197
4	The Multi-class Ensemble Framework	199
4.1	Ensemble of Multi-class Submodels	199
4.2	Hierarchical Separate-and-Conquer Ensemble	200
4.3	One-versus-Rest Ensemble	201
4.4	An Illustrative Example (Cont'd)	202
5	Experiments	203
5.1	Binary Classification Problems	205
5.2	Multi-class Classification Problems	206
5.3	Comparison of the Ensemble Methods	206
5.4	Comparison of Different Feature Selection Methods	208
6	Conclusions	210
	References	210
	Independent Data Model Selection for Ensemble Dispersion Forecasting	213
	<i>Angelo Ciaramella, Giulio Giunta, Angelo Riccio, Stefano Galmarini</i>	
1	Introduction	213
2	The 'Median Model' Approach	215
3	Negentropy-Based Hierarchical Agglomeration	218
3.1	Kullback-Leibler Divergence	218
3.2	Negentropy Information	218
3.3	Agglomerative Approach	219
4	Experimental Results	219
5	Conclusions	229
	References	229
	Integrating Liknon Feature Selection and Committee Training	233
	<i>Erinija Pranckeviciene</i>	
1	Introduction	233
2	Computational Paradigm and Parameters	234
2.1	Liknon-Based Feature Selection	235
2.2	Banana Example: Classification in Nonlinear Class Separation	238
3	The Benchmark of NIPS2003 Feature Selection Challenge	240
3.1	Performance, Size and Purity of the Feature Subsets in the Benchmark	240

4	NN3 Committee and Liknon Feature Profiles in the Benchmark	242
5	Conclusion	248
	References	248
Evaluating Hybrid Ensembles for Intelligent Decision Support for Intensive Care		
<i>Pedro Gago, Manuel Filipe Santos</i>		
1	Introduction	251
2	Previous Work	252
3	INTCare System	254
	3.1 Description of the Agents	254
4	Problem Description	256
	4.1 Data Description	257
5	Related Work	257
6	Experimental Setting	258
7	Results	260
8	Discussion	262
9	Conclusion	263
	References	263
Index		267