

Contents

1	Facial Action Unit Recognition Using Filtered Local Binary Pattern Features with Bootstrapped and Weighted ECOC Classifiers	1
	<i>Raymond S. Smith, Terry Windeatt</i>	
1.1	Introduction	1
1.2	Theoretical Background	5
1.2.1	ECOC Weighted Decoding	5
1.2.2	Platt Scaling	6
1.2.3	Local Binary Patterns	7
1.2.4	Fast Correlation-Based Filtering	8
1.2.5	Principal Components Analysis	9
1.3	Algorithms	10
1.4	Experimental Evaluation	10
1.4.1	Classifier Accuracy	13
1.4.2	The Effect of Platt Scaling	14
1.4.3	A Bias/Variance Analysis	15
1.5	Conclusion	16
1.6	Code Listings	17
	References	19
2	On the Design of Low Redundancy Error-Correcting Output Codes	21
	<i>Miguel Ángel Bautista, Sergio Escalera, Xavier Baró, Oriol Pujol, Jordi Vitrià, Petia Radeva</i>	
2.1	Introduction	21
2.2	Compact Error-Correcting Output Codes	23
2.2.1	Error-Correcting Output Codes	23
2.2.2	Compact ECOC Coding	24
2.3	Results	29
2.3.1	UCI Categorization	30
2.3.2	Computer Vision Applications	32

2.4	Conclusion	36
	References	37
3	Minimally-Sized Balanced Decomposition Schemes for Multi-class Classification	39
	<i>Evgueni N. Smirnov, Matthijs Moed, Georgi Nalbantov, Ida Sprinkhuizen-Kuyper</i>	
3.1	Introduction	40
3.2	Classification Problem	41
3.3	Decomposing Multi-class Classification Problems	41
3.3.1	Decomposition Schemes	41
3.3.2	Encoding and Decoding	44
3.4	Balanced Decomposition Schemes and Their Minimally-Sized Variant	46
3.4.1	Balanced Decomposition Schemes	46
3.4.2	Minimally-Sized Balanced Decomposition Schemes	47
3.4.3	Voting Using Minimally-Sized Balanced Decomposition Schemes	49
3.5	Experiments	51
3.5.1	UCI Data Experiments	51
3.5.2	Experiments on Data Sets with Large Number of Classes	52
3.5.3	Bias-Variance Decomposition Experiments	54
3.6	Conclusion	55
	References	56
4	Bias-Variance Analysis of ECOC and Bagging Using Neural Nets ...	59
	<i>Cemre Zor, Terry Windeatt, Berrin Yanikoglu</i>	
4.1	Introduction	59
4.1.1	Bootstrap Aggregating (Bagging)	60
4.1.2	Error Correcting Output Coding (ECOC)	60
4.1.3	Bias and Variance Analysis	62
4.2	Bias and Variance Analysis of James	64
4.3	Experiments	65
4.3.1	Setup	65
4.3.2	Results	68
4.4	Discussion	72
	References	72
5	Fast-Ensembles of Minimum Redundancy Feature Selection	75
	<i>Benjamin Schowe, Katharina Morik</i>	
5.1	Introduction	75
5.2	Related Work	76
5.2.1	Ensemble Methods	78
5.3	Speeding Up Ensembles	78
5.3.1	Inner Ensemble	79

5.3.2	Fast Ensemble	80
5.3.3	Result Combination	84
5.3.4	Benefits	85
5.4	Evaluation	85
5.4.1	Stability	86
5.4.2	Accuracy	87
5.4.3	Runtime	92
5.4.4	LUCAS	93
5.5	Conclusion	94
	References	95
6	Hybrid Correlation and Causal Feature Selection for Ensemble Classifiers	97
	<i>Rakkrit Duangsoithong, Terry Windeatt</i>	
6.1	Introduction	97
6.2	Related Research	99
6.3	Theoretical Approach	100
6.3.1	Feature Selection Algorithms	100
6.3.2	Causal Discovery Algorithm	102
6.3.3	Feature Selection Analysis	103
6.3.4	Ensemble Classifier	106
6.3.5	Pseudo-code: Hybrid Correlation and Causal Feature Selection for Ensemble Classifiers Algorithm	106
6.4	Experimental Setup	108
6.4.1	Dataset	108
6.4.2	Evaluation	109
6.5	Experimental Result	110
6.6	Discussion	113
6.7	Conclusion	114
	References	114
7	Learning Markov Blankets for Continuous or Discrete Networks via Feature Selection	117
	<i>Houtao Deng, Saylisse Davila, George Runger, Eugene Tuv</i>	
7.1	Introduction	117
7.1.1	Learning Bayesian Networks Via Feature Selection	118
7.2	Feature Selection Framework	119
7.2.1	Feature Importance Measure	120
7.2.2	Feature Masking Measure and Its Relationship to Markov Blanket	121
7.2.3	Statistical Criteria for Identifying Relevant and Redundant Features	124
7.2.4	Residuals for Multiple Iterations	124
7.3	Experiments	125
7.3.1	Continuous Gaussian Local Structure Learning	125
7.3.2	Continuous Non-Gaussian Local Structure Learning	127

7.3.3	Discrete Local Structure Learning	128
7.4	Conclusion	130
	References	130
8	Ensembles of Bayesian Network Classifiers Using Glaucoma Data and Expertise	133
	<i>Stefano Ceccon, David Garway-Heath, David Crabb, Allan Tucker</i>	
8.1	Improving Knowledge and Classification of Glaucoma	133
8.2	Theory and Methods	134
8.2.1	Datasets	134
8.2.2	Bayesian Networks	135
8.2.3	Combining Networks	140
8.3	Algorithms	141
8.3.1	Learning the Structure	141
8.3.2	Combining Two Networks	142
8.3.3	Optimized Combination	143
8.4	Results and Performance Evaluation	143
8.4.1	Base Classifiers	143
8.4.2	Ensembles of Classifiers	144
	References	148
9	A Novel Ensemble Technique for Protein Subcellular Location Prediction	151
	<i>Alessandro Rozza, Gabriele Lombardi, Matteo Re, Elena Casiraghi, Giorgio Valentini, Paola Campadelli</i>	
9.1	Introduction	151
9.2	Related Works	153
9.3	Classifiers Based on Efficient Fisher Subspace Estimation	156
9.3.1	A Kernel Version of TIPCAC	157
9.4	DDAG K-TIPCAC	158
9.4.1	Decision DAGs (DDAGs)	158
9.4.2	Decision DAG K-TIPCAC	158
9.5	Experimental Setting	159
9.5.1	Methods	159
9.5.2	Dataset	160
9.5.3	Performance Evaluation	161
9.6	Results	161
9.6.1	DDAG K-TIPCAC Employing the Standard Multiclass Estimation of F_s	163
9.6.2	DDAG K-TIPCAC without Projection on Multiclass F_s ..	164
9.7	Conclusion	165
	References	166

10	Trading-Off Diversity and Accuracy for Optimal Ensemble Tree Selection in Random Forests	169
	<i>Haytham Elghazel, Alex Aussem, Florence Perraud</i>	
10.1	Introduction	169
10.2	Background of Ensemble Selection	171
10.3	Contribution	172
10.4	Empirical Results	174
10.4.1	Experiments on Benchmark Data Sets	174
10.4.2	Experiments on Real Data Sets	175
10.5	Conclusion	177
	References	178
11	Random Oracles for Regression Ensembles	181
	<i>Carlos Pardo, Juan J. Rodríguez, José F. Díez-Pastor, César García-Osorio</i>	
11.1	Introduction	181
11.2	Random Oracles	183
11.3	Experiments	183
11.4	Results	185
11.5	Diversity-Error Diagrams	191
11.6	Conclusion	194
	References	198
12	Embedding Random Projections in Regularized Gradient Boosting Machines	201
	<i>Pierluigi Casale, Oriol Pujol, Petia Radeva</i>	
12.1	Introduction	201
12.2	Related Works on RPs	202
12.3	Methods	203
12.3.1	Gradient Boosting Machines	203
12.3.2	Random Projections	204
12.3.3	Random Projections in Boosting Machine	205
12.4	Experiments and Results	206
12.4.1	Test Patterns	207
12.4.2	UCI Datasets	209
12.4.3	The Effect of Regularization in RpBoost	211
12.4.4	Discussion	214
12.5	Conclusion	215
	References	216
13	An Improved Mixture of Experts Model: Divide and Conquer Using Random Prototypes	217
	<i>Giuliano Armano, Nima Hatami</i>	
13.1	Introduction	217
13.2	Standard Mixture of Experts Models	220
13.2.1	Standard ME Model	220

13.2.2	Standard HME Model	221
13.3	Mixture of Random Prototype-Based Experts (MRPE) and Hierarchical MRPE	222
13.3.1	Mixture of Random Prototype-Based Local Experts	222
13.3.2	Hierarchical MRPE Model	225
13.4	Experimental Results and Discussion	227
13.5	Conclusion	230
	References	230
14	Three Data Partitioning Strategies for Building Local Classifiers . . .	233
	<i>Indrė Žliobaite</i>	
14.1	Introduction	233
14.2	Three Alternatives for Building Local Classifiers	234
14.2.1	Instance Based Partitioning	235
14.2.2	Instance Based Partitioning with Label Information	236
14.2.3	Partitioning Using One Feature	236
14.3	Analysis with the Modeling Dataset	238
14.3.1	Testing Scenario	239
14.3.2	Results	242
14.4	Experiments with Real Data	242
14.4.1	Datasets	242
14.4.2	Implementation Details	243
14.4.3	Experimental Goals	243
14.4.4	Results	244
14.5	Conclusion	249
	References	250
	Index	251