

Contents

1	Advances in Feature Selection for Data and Pattern Recognition: An Introduction	1
	Urszula Stańczyk, Beata Zielosko and Lakhmi C. Jain	
1.1	Introduction	1
1.2	Chapters of the Book	3
1.3	Concluding Remarks	7
	References	7
 Part I Nature and Representation of Data		
2	Attribute Selection Based on Reduction of Numerical Attributes During Discretization	13
	Jerzy W. Grzymała-Busse and Teresa Mroczek	
2.1	Introduction	13
2.2	Dominant Attribute Discretization	14
2.3	Multiple Scanning Discretization	19
2.4	Experiments	19
2.5	Conclusions	23
	References	23
3	Improving Bagging Ensembles for Class Imbalanced Data by Active Learning	25
	Jerzy Błaszczyński and Jerzy Stefanowski	
3.1	Introduction	25
3.2	Improving Classifiers Learned from Imbalanced Data	28
3.2.1	Nature of Imbalanced Data	28
3.2.2	Evaluation of Classifiers on Imbalanced Data	29
3.2.3	Main Approaches to Improve Classifiers for Imbalanced Data	31
3.3	Active Learning	32

3.4	Ensembles Specialized for Imbalanced Data	33
3.5	Active Selection of Examples in Under-Sampling Bagging	35
3.6	Experimental Evaluation	39
3.6.1	Experimental Setup	39
3.6.2	Results of Experiments	40
3.7	Conclusions	48
	References	50
4	Attribute-Based Decision Graphs and Their Roles in Machine Learning Related Tasks	53
	João Roberto Bertini Junior and Maria do Carmo Nicoletti	
4.1	Introduction	54
4.2	The Attribute-Based Decision Graph Structure (AbDG) for Representing Training Data	56
4.2.1	Constructing an AbDG	56
4.2.2	Assigning Weights to Vertices and Edges	58
4.2.3	Computational Complexity for Building an AbDG	61
4.3	Using the AbDG Structure for Classification Tasks	62
4.4	Using an AbDG for Classification Purposes - A Case Study . .	63
4.5	Using the AbDG Structure for Imputation Tasks	64
4.6	Searching for Refined AbDG Structures via Genetic Algorithms	66
4.7	Conclusions	69
	References	70
5	Optimization of Decision Rules Relative to Length Based on Modified Dynamic Programming Approach	73
	Beata Zielosko and Krzysztof Żabiński	
5.1	Introduction	73
5.2	Background	75
5.3	Main Notions	76
5.4	Modified Algorithm for Directed Acyclic Graph Construction $\Delta_{\beta}^*(T)$	78
5.5	Procedure of Optimization Relative to Length	81
5.6	Experimental Results	83
5.6.1	Attributes' Values Selection and Size of the Graph . . .	83
5.6.2	Comparison of Length of β -Decision Rules	85
5.6.3	Classifier Based on Rules Optimized Relative to Length	90
5.7	Conclusions	91
	References	91

Part II Ranking and Exploration of Features

6	Generational Feature Elimination and Some Other Ranking Feature Selection Methods	97
	Wiesław Paja, Krzysztof Pancerz and Piotr Grochowalski	
6.1	Introduction	98
6.2	Selected Methods and Algorithms	99
6.2.1	Rough Set Based Feature Selection	100
6.2.2	Random Forest Based Feature Selection	102
6.2.3	Generational Feature Elimination Algorithm	103
6.3	Feature Selection Experiments	105
6.4	Results and Conclusions	107
	References	110
7	Ranking-Based Rule Classifier Optimisation	113
	Urszula Stańczyk	
7.1	Introduction	113
7.2	Background	114
7.2.1	Attribute Ranking	115
7.2.2	Rule Classifiers	116
7.2.3	Filtering Rules	117
7.3	Research Framework	118
7.3.1	Preparation of the Input Datasets	118
7.3.2	Rankings of Features	119
7.3.3	DRSA Decision Rules	120
7.3.4	Weighting Rules	122
7.4	Experimental Results	123
7.4.1	Filtering Rules by Attributes	123
7.4.2	Filtering Rules by Weights	126
7.4.3	Summary of Results	127
7.5	Conclusions	129
	References	130
8	Attribute Selection in a Dispersed Decision-Making System	133
	Małgorzata Przybyła-Kasperek	
8.1	Introduction	133
8.2	Related Works	134
8.3	Basics of the Rough Set Theory	135
8.4	An Overview of Dispersed Systems	136
8.4.1	Basic Definitions	137
8.4.2	Static Structure	138
8.4.3	Dynamic Structure with Disjoint Clusters	139
8.4.4	Dynamic Structure with Inseparable Clusters	141
8.4.5	Dynamic Structure with Negotiations	141

8.5	Description of the Experiments	143
8.5.1	Data Sets	143
8.5.2	Attribute Selection	144
8.5.3	Evaluation Measures and Parameters Optimization	145
8.6	Experiments and Discussion	148
8.6.1	Results for Static Structure	148
8.6.2	Results for Dynamic Structure with Disjoint Clusters	157
8.6.3	Results for Dynamic Structure with Inseparable Clusters	157
8.6.4	Results for Dynamic Structure with Negotiations	157
8.6.5	Comparison of All Methods	158
8.6.6	Comparison for Data Sets	160
8.7	Conclusion	160
	References	161
9	Feature Selection Approach for Rule-Based Knowledge Bases	163
	Agnieszka Nowak-Brzezińska	
9.1	Introduction	163
9.2	Feature Selection Methods for Rule Mining Processes	164
9.2.1	Significance of Rules Mining Process	166
9.2.2	Feature Selection in the Rules and Their Clusters	167
9.2.3	Related Works	168
9.2.4	Clustering the Rules Based on the Similarity Approach	168
9.2.5	Clustering the Rules	170
9.2.6	Cluster's Representative	172
9.3	Rough Set Approach in Creating Rules' Clusters Representatives	173
9.3.1	Lower and Upper Approximation Approach	174
9.3.2	KbExplorer - A Tool for Knowledge Base Exploration	175
9.4	Experiments	176
9.4.1	From Original Data to the Rules	176
9.4.2	The Results of the Experiments	176
9.4.3	The Summary of the Experiments	180
9.5	Conclusion	180
	References	181
 Part III Image, Shape, Motion, and Audio Detection and Recognition		
10	Feature Selection with a Genetic Algorithm for Classification of Brain Imaging Data	185
	Annamária Szenkovits, Regina Meszlényi, Krisztian Buza, Noémi Gaskó, Rodica Ioana Lung and Mihai Suciu	

10.1	Introduction	186
10.2	Materials and Methods	189
10.2.1	Data and Preprocessing	189
10.2.2	Least Absolute Shrinkage and Selection Operator	189
10.2.3	Minimalist Genetic Algorithm for Feature Selection	191
10.2.4	Experimental Set-Up	194
10.3	Results	195
10.4	Discussion	198
10.5	Conclusions	199
	References	200
11	Shape Descriptions and Classes of Shapes. A Proximal Physical Geometry Approach	203
	James Francis Peters and Sheela Ramanna	
11.1	Introduction	204
11.2	Introduction to Simplicial Complexes	205
11.2.1	Examples: Detecting Shapes from Simplicial Complexes	206
11.3	Preliminaries of Proximal Physical Geometry	209
11.3.1	Image Object Shape Geometry	209
11.3.2	Descriptions and Proximities	210
11.3.3	Descriptive Proximities	212
11.3.4	Edelsbrunner–Harer Nerve Simplexes	215
11.4	Features of Image Object Shapes	218
11.5	Concluding Remarks	223
	References	223
12	Comparison of Classification Methods for EEG Signals of Real and Imaginary Motion	227
	Piotr Szczuko, Michał Lech and Andrzej Czyżewski	
12.1	Introduction	228
12.2	EEG Signal Parameterisation	229
12.3	Data Classification Method	232
12.4	Classification Results	234
12.5	Conclusions	236
	References	237
13	Application of Tolerance Near Sets to Audio Signal Classification	241
	Ashmeet Singh and Sheela Ramanna	
13.1	Introduction	241
13.2	Related Work	243
13.3	Automatic Detection of Video Blocks	245
13.4	Audio and Video Features	246
13.4.1	Audio Features	246
13.4.2	Video Features	249

13.5	Theoretical Framework: Near Sets and Tolerance Near Sets . . .	249
13.5.1	Preliminaries	250
13.5.2	Tolerance Near Sets	252
13.6	Tolerance Class Learner - TCL	254
13.6.1	Algorithms	255
13.6.2	Phase II: Classification	257
13.7	Experiments	257
13.7.1	Speech and Non-speech Dataset	257
13.7.2	Discussion	259
13.7.3	Music and Music + Speech Dataset	260
13.7.4	Detection of Commercial Blocks in News Data	261
13.8	Conclusion and Future Work	263
	References	263

Part IV Decision Support Systems

14	Visual Analysis of Relevant Features in Customer Loyalty Improvement Recommendation	269
	Katarzyna A. Tarnowska, Zbigniew W. Raś, Lynn Daniel and Doug Fowler	
14.1	Introduction	270
14.2	Related Applications	270
14.3	Dataset and Application	271
14.3.1	Dataset	271
14.4	Decision Problem	272
14.4.1	Attribute Analysis	272
14.4.2	Attribute Reduction	273
14.4.3	Customer Satisfaction Analysis and Recognition	273
14.4.4	Providing Recommendations	274
14.5	Proposed Approach	274
14.5.1	Machine Learning Techniques	274
14.5.2	Visualization Techniques	282
14.6	Evaluation Results	286
14.6.1	Single Client Data (Local) Analysis	287
14.6.2	Global Customer Sentiment Analysis and Prediction	288
14.6.3	Recommendations for a Single Client	289
14.7	Conclusions	292
	References	292
15	Evolutionary and Aggressive Sampling for Pattern Revelation and Precognition in Building Energy Managing System with Nature-Based Methods for Energy Optimization	295
	Jarosław Utracki and Mariusz Boryczka	
15.1	Introduction	296

15.2	Ant Colony Optimization	296
15.2.1	Ant System	296
15.2.2	Ant Colony Optimization Metaheuristics	298
15.2.3	Ant Colony System	299
15.2.4	Type of Issues That Are Able To Be Solved Using <i>ACO</i>	302
15.3	Building’s Installation as 3D Matrix-Like Grids	303
15.4	Classical <i>ACO</i> Approach in Revealing a Pattern	306
15.4.1	Capability of Revealing with <i>ACO</i>	306
15.5	Aggressive <i>ACO</i> Sampling in Revealing a Pattern	308
15.5.1	Ant Decomposition and Specialization	308
15.5.2	Precognition Paths	309
15.5.3	Ants Spreading	311
15.6	Algorithm and Procedures	312
15.7	Conclusion	317
	References	318
	Index	321

Advances in Feature Selection for Data and Pattern
Recognition

Stanczyk, U.; Zielosko, B.; Jain, L.C. (Eds.)

2018, XVIII, 328 p. 37 illus., 20 illus. in color., Hardcover

ISBN: 978-3-319-67587-9