

Table of Contents

1. Evolutionary Algorithms for Data Mining and Knowledge Discovery

Ashish Ghosh	1
1.1 Introduction	1
1.2 Knowledge Discovery in Databases.....	2
1.2.1 Data Preprocessing	2
1.2.2 Data Mining	3
1.2.3 Knowledge Interpretation or Postprocessing.....	5
1.2.4 Desirable Properties of Discovered Knowledge.....	6
1.2.5 Applications	6
1.3 Genetic Algorithms (GAs) for Rule Discovery	9
1.3.1 Association Rule Mining Algorithms: An Overview ...	9
1.3.2 Individual Representation	10
1.3.3 Representing the Rule Antecedent (a Conjunction of Conditions)	11
1.3.4 Representing the Rule Consequent (Predicted Class) ..	12
1.3.5 Genetic Operators for Rule Discovery	12
1.3.6 Fitness Functions for Rule Discovery	14
1.4 Multi-Objective Optimization and Rule Mining Problems ...	15
1.4.1 Model of Ghosh and Nath	16
1.5 Discussion and Research Directions	18
References	19

2. Strategies for Scaling Up Evolutionary Instance Reduction Algorithms for Data Mining

Jose Ramon Cano, Francisco Herrera, and Manuel Lozano.....	21
2.1 Introduction	21
2.2 Instance Selection on Data Reduction	23
2.2.1 Instance Selection	23
2.2.2 Instance Selection for Prototype Selection	23
2.2.3 Overview of Instance Selection Algorithms.....	24
2.3 Evolutionary Instance Selection Algorithms	25
2.3.1 Evolutionary Instance Selection: Key Points	26
2.3.2 The CHC Algorithm	26

2.4	The Scaling up Problem. The Stratified Approach	27
2.4.1	Scaling up and Stratification	27
2.4.2	Evolutionary Algorithms and Stratification Strategy	28
2.5	Experimental Methodology	29
2.5.1	Data Sets and Parameters	29
2.5.2	Partitions and Stratification: A Specific Model	30
2.5.3	Table of Results	32
2.6	Experimental Study	32
2.6.1	Medium Size Data Sets	32
2.6.2	Large Size Data Set	36
2.7	Concluding Remarks	38
	References	38
 3. GAP: Constructing and Selecting Features with Evolutionary Computing		
	Matthew G. Smith and Larry Bull	41
3.1	Introduction	41
3.2	The GAP Algorithm	43
3.2.1	Feature Creation	43
3.2.2	Feature Selection	45
3.3	Experimentation	47
3.3.1	Results	48
3.3.2	Analysis	49
3.3.3	The Importance of Reordering the Dataset	51
3.3.4	Combining Creation and Selection in a Single Stage...	52
3.3.5	A Rough Comparison to Other Algorithms	54
3.4	Conclusion	54
	References	55
 4. Multi-Agent Data Mining using Evolutionary Computing		
	Riyaz Sikora	57
4.1	Introduction	57
4.2	Related Work	58
4.2.1	Feature Selection	58
4.2.2	Distributed Data Mining	60
4.3	Data Mining with GA	60
4.4	Feature Selection as Self-Adaptation	62
4.5	Multi-Agent Data Mining	63
4.6	Experimental Results	66
4.6.1	Experimental Design	66
4.6.2	Effectiveness of Feature Selection	66
4.6.3	Effectiveness of DLS	69
4.7	Conclusions and Future Work	75
	References	76

5. A Rule Extraction System with Class-Dependent Features

Xiuju Fu and Lipo Wang	79
5.1 Overview	79
5.1.1 Rule Extraction Systems	79
5.1.2 Categories of Rule Extraction Systems	81
5.1.3 Data Dimensionality Reduction	82
5.2 Our Rule Extraction System	87
5.3 RBF Classifiers	88
5.3.1 A Conventional RBF Classifier	88
5.3.2 A Novel RBF Classifier	89
5.3.3 Feature Masks Encoded by GA	91
5.4 Rule Extraction by the Gradient Descent Method	92
5.5 Experimental Results	95
5.5.1 Thyroid Data Set	95
5.5.2 Wine Data Set	96
5.5.3 Summary	96
References	97

6. Knowledge Discovery in Data Mining via an Evolutionary Algorithm

Qi Yu, Kay Chen Tan, and Tong Heng Lee	101
6.1 Introduction	101
6.2 Coevolutionary Rule Extractor	103
6.2.1 Coevolutionary Algorithms	103
6.2.2 Overall Flowchart of CORE	104
6.2.3 Population and Chromosome Structure	106
6.2.4 Genetic Operators	107
6.2.5 Fitness Evaluations	107
6.2.6 Token Competition	108
6.3 Case Study	109
6.3.1 Experimental Setup	109
6.3.2 Simulation Results and Comparisons	109
6.3.3 Discussion and Summary	118
6.4 Conclusions	121
References	121

7. Diversity and Neuro-Ensemble

Minh Ha Nguyen, Hussein Abbass, and Robert McKay	125
7.1 Introduction	125
7.2 Ensemble of Predictors	126
7.2.1 Forming and Selecting the Individual Networks	126
7.2.2 Combining Neural Networks to form the Ensemble ...	128
7.2.3 Review of Diversity Promotion Mechanisms	129

7.2.4	Review of the use of Diversity in Selecting the Members for the Ensemble	132
7.3	Comparisons of Different Ensemble Methods	134
7.3.1	Methods	134
7.3.2	Evolutionary Multi Objective Optimization Application in Ensemble	136
7.3.3	Experiment Set-up	138
7.3.4	Results and Analysis	141
7.4	Conclusion	150
	References	153

8. Unsupervised Niche Clustering:

Discovering an Unknown Number of Clusters in Noisy

Data Sets

Olfa Nasraoui, Elizabeth Leon, and Raghu Krishnapuram	157
---	-----

8.1	Introduction	157
8.1.1	Unsupervised Clustering	158
8.1.2	Motivations for Evolutionary Clustering	158
8.1.3	Why Genetic Niching?	159
8.2	Genetic Niching	160
8.3	An Overview of Existing Evolutionary Clustering Techniques	161
8.4	New Approach to Unsupervised Robust Clustering using Genetic Niching	163
8.4.1	Representation	163
8.4.2	Fitness Function	164
8.4.3	Analogy between Density and Scale in Artificial Niches and Fertility and Barrier Constraints in Natural Niches	166
8.4.4	A Baldwin Effect for Scale Estimation	166
8.4.5	Mating Restriction	168
8.4.6	Scale Inheritance	169
8.4.7	Crossover and Mutation	169
8.4.8	Incorporating Constraints Against Degenerate and Spurious Solutions	169
8.4.9	Selecting the Initial Population	171
8.4.10	Extracting Cluster Centers From the Final Population	171
8.4.11	Refinement of the Extracted Prototypes	172
8.4.12	The Unsupervised Niche Clustering Algorithm (UNC)	173
8.4.13	Computational Complexity	173
8.5	Simulation Results on Synthetic Examples	174
8.5.1	Detailed Phases of Cluster Evolution	174
8.5.2	Sensitivity to GA Parameters, Noise, and Effect of Final Refinement	174

8.5.3	Simulation Results for Data with Varying Number of Clusters, Cluster Sizes, Densities, and Noise Contamination Rates	178
8.6	Application to Image Segmentation	179
8.7	Conclusion	185
	References	186
9.	Evolutionary Computation in Intelligent Network Management	
	Ajith Abraham	189
9.1	Intrusion Detection Systems	189
9.1.1	Intrusion Detection - a Data Mining Approach	191
9.1.2	Linear Genetic Programming (LGP)	192
9.1.3	Decision Trees (DT) as Intrusion Detection Model	192
9.1.4	Support Vector Machines (SVM)	193
9.1.5	Intrusion Detection Data	193
9.1.6	Discussions	198
9.2	Web usage Mining using Intelligent Miner (i-Miner)	198
9.2.1	Optimization of Fuzzy Clustering Algorithm	199
9.2.2	Optimization of the Fuzzy Inference System	200
9.2.3	Discussions	206
9.3	Conclusions	206
	References	209
10.	Genetic Programming in Data Mining for Drug Discovery	
	W. B. Langdon and S. J. Barrett	211
10.1	Computational Drug Discovery	213
10.2	Evolutionary Computing for Drug Discovery	214
10.3	Oral Bioavailability	214
10.4	Genetic Programming	215
10.5	Receiver Operating Characteristics	215
10.5.1	Simple use of ROC as the Objective to be Maximized .	216
10.6	The Bioavailability Data	221
10.7	Chemical Features	221
10.8	Genetic Programming Configuration	223
10.8.1	Function set	223
10.8.2	Terminal set	223
10.8.3	GP Genetic Operations and other Parameters	223
10.8.4	GP Fitness Function	225
10.9	Experiments	225
10.9.1	Training on 321 Human Records	226
10.9.2	Training on 1342 Rat Records	226
10.9.3	Simplification of Evolved Model of Rat Bioavailability .	228

10.10 Discussion	231
10.11 Conclusion	232
References	232
11. Microarray Data Mining with Evolutionary Computation	
Gary B. Fogel	237
11.1 Introduction	237
11.2 Microarrays	238
11.3 Supervised and Unsupervised Methods for Microarray Analysis	239
11.3.1 Supervised Methods of Class Assignment	239
11.4 Evolutionary Computation	241
11.4.1 Evolutionary Computation in Bioinformatics	242
11.5 Applications of Evolutionary Computation for Class Prediction and Expression Clustering	242
11.6 Conclusions	245
References	245
12. An Evolutionary Modularized Data Mining Mechanism for Financial Distress Forecasts	
Po-Chang (P.C.) Ko and Ping-Chen (P.C.) Lin	249
12.1 Introduction	249
12.2 Evolutionary Modularized Mining Mechanism	250
12.2.1 Exploration Process	251
12.2.2 Transformation Process	252
12.2.3 Mining Process	252
12.3 Implementations	253
12.4 Performance Evaluation	256
12.4.1 Exploration Process	256
12.4.2 Mining Process	257
12.4.3 Simulation Results	258
12.5 Conclusions	261
References	262
Subject Index	264



<http://www.springer.com/978-3-540-22370-2>

Evolutionary Computation in Data Mining

Ghosh, A. (Ed.)

2005, XVIII, 266 p., Hardcover

ISBN: 978-3-540-22370-2