

Contents

<i>Preface</i>	vii
1. Knowledge Discovery and Data Mining: Concepts and Fundamental Aspects	1
1.1 Overview	1
1.2 Data Mining and Knowledge Discovery	1
1.3 Taxonomy of Data Mining Methods	3
1.4 Supervised Methods	4
1.4.1 Overview	4
1.4.2 Training Set	5
1.4.3 Definition of the Classification Problem	6
1.4.4 Induction Algorithms	8
1.5 Rule Induction	9
1.6 Decision Trees	10
1.7 Bayesian Methods	12
1.7.1 Overview	12
1.7.2 Naïve Bayes	12
1.7.2.1 The Basic Naïve Bayes Classifier	12
1.7.2.2 Naïve Bayes for Numeric Attributes	14
1.7.2.3 Correction to the Probability Estimation	14
1.7.2.4 Laplace Correction	15
1.7.2.5 No Match	15
1.7.3 Other Bayesian Methods	16
1.8 Other Induction Methods	16
1.8.1 Neural Networks	16
1.8.2 Genetic Algorithms	18

1.8.3	Instancebased Learning	19
1.8.4	Support Vector Machines	19
1.9	Performance Evaluation	20
1.9.1	Generalization Error	20
1.9.2	Theoretical Estimation of Generalization Error	21
1.9.2.1	VC-Framework	21
1.9.2.2	PAC-Framework	23
1.9.3	Empirical Estimation of Generalization Error	23
1.9.4	Bias and Variance Decomposition	25
1.9.5	Computational Complexity	26
1.9.6	Comprehensibility	27
1.10	“No Free Lunch” Theorem	28
1.11	Scalability to Large Datasets	30
1.12	The “Curse of Dimensionality”	32
2.	Decision Trees	35
2.1	Decision Trees	35
2.2	Algorithmic Framework for Decision Trees	37
2.3	Univariate Splitting Criteria	38
2.3.1	Overview	38
2.3.2	Impurity Based Criteria	38
2.3.3	Information Gain	40
2.3.4	Gini Index	40
2.3.5	Likelihood Ratio Chi-Squared Statistics	40
2.3.6	DKM Criterion	41
2.3.7	Normalized Impurity Based Criteria	41
2.3.8	Gain Ratio	41
2.3.9	Distance Measure	42
2.3.10	Binary criteria	42
2.3.11	Twoing Criterion	42
2.3.12	Orthogonal Criterion	43
2.3.13	Kolmogorov–Smirnov Criterion	43
2.3.14	AUC Splitting Criteria	44
2.3.15	Other Univariate Splitting Criteria	44
2.3.16	Comparison of Univariate Splitting Criteria	44
2.4	Multivariate Splitting Criteria	44
2.5	Stopping Criteria	45
2.6	Pruning Methods	45
2.6.1	Overview	45

2.6.2	Cost-Complexity Pruning	46
2.6.3	Reduced Error Pruning	46
2.6.4	Minimum Error Pruning (MEP)	47
2.6.5	Pessimistic Pruning	47
2.6.6	Error-Based Pruning (EBP)	48
2.6.7	Optimal Pruning	49
2.6.8	Minimum Description Length Pruning	49
2.6.9	Other Pruning Methods	50
2.6.10	Comparison of Pruning Methods	50
2.7	Other Issues	50
2.7.1	Weighting Instances	50
2.7.2	Misclassification costs	50
2.7.3	Handling Missing Values	50
2.8	Decision Trees Inducers	52
2.8.1	ID3	52
2.8.2	C4.5	52
2.8.3	CART	53
2.8.4	CHAID	53
2.8.5	QUEST	54
2.8.6	Reference to Other Algorithms	54
2.8.7	Advantages and Disadvantages of Decision Trees	54
2.8.8	Oblivious Decision Trees	56
2.8.9	Decision Trees Inducers for Large Datasets	57
2.8.10	Incremental Induction	59
3.	Clustering Methods	61
3.1	Introduction	61
3.2	Distance Measures	62
3.2.1	Minkowski: Distance Measures for Numeric Attributes	62
3.2.2	Distance Measures for Binary Attributes	63
3.2.3	Distance Measures for Nominal Attributes	64
3.2.4	Distance Metrics for Ordinal Attributes	64
3.2.5	Distance Metrics for Mixed-Type Attributes	64
3.3	Similarity Functions	65
3.3.1	Cosine Measure	65
3.3.2	Pearson Correlation Measure	66
3.3.3	Extended Jaccard Measure	66
3.3.4	Dice Coefficient Measure	66

3.4	Evaluation Criteria Measures	66
3.4.1	Internal Quality Criteria	66
3.4.1.1	Sum of Squared Error (SSE)	67
3.4.1.2	Other Minimum Variance Criteria	67
3.4.1.3	Scatter Criteria	68
3.4.1.4	Condorcet's Criterion	69
3.4.1.5	The C-Criterion	70
3.4.1.6	Category Utility Metric	70
3.4.1.7	Edge Cut Metrics	70
3.4.2	External Quality Criteria	70
3.4.2.1	Mutual Information Based Measure	71
3.4.2.2	Precision-Recall Measure	71
3.4.2.3	Rand Index	71
3.5	Clustering Methods	72
3.5.1	Hierarchical Methods	72
3.5.2	Partitioning Methods	75
3.5.2.1	Error Minimization Algorithms	75
3.5.2.2	Graph-Theoretic Clustering	77
3.5.3	Density Based Methods	78
3.5.4	Model Based Clustering	80
3.5.4.1	Decision Trees	80
3.5.4.2	Neural Networks	80
3.5.5	Grid-Based Methods	81
3.5.6	Fuzzy Clustering	81
3.5.7	Evolutionary Approaches for Clustering	81
3.5.8	Simulated Annealing for Clustering	84
3.5.9	Which Technique to Use?	84
3.6	Clustering Large Data Sets	86
3.6.1	Decomposition Approach	88
3.6.2	Incremental Clustering	88
3.6.3	Parallel Implementation	90
3.7	Determining the Number of Clusters	90
3.7.1	Methods Based on Intra Cluster Scatter	91
3.7.2	Methods Based on Both the Inter and Intra Cluster Scatter	92
3.7.3	Criteria Based on Probabilistic	94
4.	Ensemble Methods	95
4.1	Multiple Classifiers	95

4.2	Ensemble Methodology	96
4.3	Sequential Methodology	97
4.3.1	Model Guided Instance Selection	98
4.3.1.1	Uncertainty Sampling	99
4.3.1.2	Boosting	100
4.3.1.3	Windowing	103
4.3.2	Incremental Batch Learning	105
4.4	Concurrent Methodology	105
4.4.0.1	Bagging	106
4.4.0.2	Cross-validated Committees	108
4.5	Combining Classifiers	108
4.5.1	Simple Combining Methods	108
4.5.1.1	Uniform Voting	108
4.5.1.2	Distribution Summation	109
4.5.1.3	Bayesian Combination	109
4.5.1.4	Dempster–Shafer	109
4.5.1.5	Naïve Bayes	110
4.5.1.6	Entropy Weighting	110
4.5.1.7	Density Based Weighting	110
4.5.1.8	DEA Weighting Method	111
4.5.1.9	Logarithmic Opinion Pool	111
4.5.1.10	Order Statistics	111
4.5.2	Meta Combining Methods	111
4.5.2.1	Stacking	111
4.5.2.2	Arbiter Trees	112
4.5.2.3	Combiner Trees	115
4.5.2.4	Grading	115
4.6	Ensemble Size	116
4.6.1	Selecting the Ensemble Size	116
4.6.2	Pruning Ensembles	117
4.7	Ensemble Diversity	118
4.7.1	Manipulating the Inducer	118
4.7.2	Manipulating the Training Set	119
4.7.2.1	Manipulating the Tuples	119
4.7.2.2	Manipulating the Input Feature Set	119
4.7.2.3	Manipulating the Target Attribute	121
4.7.3	Measuring the Diversity	121
5.	Elementary Decomposition Framework	123

5.1	Decomposition Framework	123
5.2	Decomposition Advantages	125
5.2.1	Increasing Classification Performance (Classification Accuracy)	125
5.2.2	Scalability to Large Databases	127
5.2.3	Increasing Comprehensibility	127
5.2.4	Modularity	127
5.2.5	Suitability for Parallel Computation	128
5.2.6	Flexibility in Techniques Selection	128
5.3	The Elementary Decomposition Methodology	128
5.3.1	Illustrative Example of Elementary Decomposition	131
5.3.1.1	Overview	131
5.3.1.2	Feature Set Decomposition	132
5.3.1.3	Sample Decomposition	133
5.3.1.4	Space Decomposition	133
5.3.1.5	Concept Aggregation	134
5.3.1.6	Function Decomposition	135
5.3.2	The Decomposer's Characteristics	136
5.3.2.1	Overview	136
5.3.2.2	The Structure Acquiring Method	137
5.3.2.3	The Mutually Exclusive Property	138
5.3.2.4	The Inducer Usage	139
5.3.2.5	Exhaustiveness	140
5.3.2.6	Combiner Usage	141
5.3.2.7	Sequentially or Concurrently	141
5.3.3	Distributed and Parallel Data Mining	142
5.3.4	The Uniqueness of the Proposed Decomposition Framework	143
5.3.5	Problems Formulation	145
5.3.5.1	Feature Set Decomposition	146
5.3.5.2	Space Decomposition	147
5.3.5.3	Sample Decomposition	148
5.3.5.4	Concept Aggregation	148
5.3.5.5	Function Decomposition	149
6.	Feature Set Decomposition	151
6.1	Overview	151
6.2	Problem Formulation	155
6.3	Definitions and Properties	157

6.4	Conditions for Complete Equivalence	160
6.5	Algorithmic Framework for Feature Set Decomposition . . .	167
6.5.1	Overview	167
6.5.2	Searching Methods	168
6.5.3	Induction Methods	170
6.5.4	Accuracy Evaluation Methods	174
6.5.4.1	The Wrapper Method	174
6.5.4.2	Conditional Entropy	174
6.5.4.3	VC-Dimension	175
6.5.5	Classification of an Unlabeled Instance	181
6.5.6	Computational Complexity	181
6.5.7	Specific Algorithms Implementations	183
6.5.7.1	DOG	183
6.5.7.2	BCW	184
6.6	Experimental Study	184
6.6.1	Overview	184
6.6.2	Algorithms Used	184
6.6.3	Data Sets Used	185
6.6.4	Metrics Measured	187
6.6.5	Comparison with Single Model Algorithms	188
6.6.6	Comparing to Ensemble Algorithms	189
6.6.7	Discussion	191
6.6.7.1	The Relation of Feature Set Decomposition Performance and Node-Sample Ratio	191
6.6.7.2	The Link Between Error Reduction and the Problem Complexity	192
6.6.8	The Suitability of the Naïve Bayes Combination . . .	193
6.6.9	The Effect of Subset Size	193
7.	Space Decomposition	201
7.0.10	Overview	201
7.0.11	Motivation	201
7.1	Problem Formulation	204
7.1.1	Manners for Dividing the Instance Space	205
7.1.2	The K -Means Algorithm as a Decomposition Tool . .	206
7.1.3	Determining the Number of Subsets	207
7.1.4	The Basic K -Classifier Algorithm	208
7.2	The Heterogeneity Detecting K -Classifier (HDK-Classifier)	210

7.2.1	Overview	210
7.2.2	Running-Time Complexity	211
7.3	Experimental Study	211
7.4	Conclusions	212
8.	Sample Decomposition	215
8.1	Overview	215
8.2	Tuple Sampling	215
8.3	Problem Formulation	216
8.4	The CBCD Algorithm	217
8.4.1	Stages of the Algorithm	218
8.4.1.1	Stage 1 — Apply the K -Means Algorithm on the Entire DataSet S	218
8.4.1.2	Stage 2 — Produce the ψ Subsets Based on the K Clusters Created Sn Stage 1.	218
8.4.1.3	Stage 3 — Produce ψ Classifiers by Apply- ing the C4.5 Learning Algorithm on the ψ Cluster-Based Subsets.	219
8.4.2	Running-Time Complexity Analysis	219
8.4.3	Classifying New Instances	220
8.5	An Illustrative Example	220
8.5.1	Step 1 — Creating a Training Set	221
8.5.2	Step 2 — Performing Clustering	222
8.5.3	Step 3 — Creating the Subsets	222
8.5.4	Step 4 — Performing Induction	222
8.5.5	Step 5 — Testing the Classifiers	223
8.6	Experimental Study	224
8.6.1	Data Sets Used	225
8.6.2	Experimental Stages	225
8.6.2.1	Stage 1 — Partitioning the Dataset	225
8.6.2.2	Stage 2 — Applying the CBCD Algorithm on the Datasets	226
8.6.2.3	Stage 3 — Employing the Paired T Test	226
8.7	Conclusions	229
9.	Function Decomposition	231
9.1	Data Preprocessing	231
9.2	Feature Transformation	232

9.2.1	Feature Aggregation	234
9.3	Supervised and Unsupervised Discretization	236
9.4	Function Decomposition	236
9.4.1	Boolean Function Decomposition in Switching Circuits	236
9.4.2	Feature Discovery by Constructive Induction	237
9.4.3	Function Decomposition in Data Mining	238
9.4.4	Problem Formulation	241
9.5	The IBAC Algorithm	241
9.5.1	The Basic Stage	243
9.5.1.1	Initialization	243
9.5.1.2	Classifier Comparison	244
9.5.2	Assigning Values to the New Attribute	244
9.5.3	Creating the New Dataset	245
9.5.4	Selection Criterion and Stopping Criterion	245
9.5.5	Time Complexity Analysis	246
9.5.6	Classifying New Instances	247
9.5.7	Example	247
9.6	Experimental Study	249
9.6.1	Databases Used	249
9.6.2	Dataset Discretization	249
9.6.3	Evaluating the Algorithm Performance	250
9.6.4	Experimental Results	250
9.7	Meta-Classifer	252
9.7.1	Building the Meta-Database	253
9.7.2	Inducing the Meta-Classifer	254
9.7.3	Examining the Meta-Learning Approach	254
9.7.4	Meta-Classifer Results	255
9.8	Conclusions	257
10.	Concept Aggregation	259
10.1	Overview	259
10.2	Motivation	260
10.3	The Multi Criteria Problem	261
10.4	Problem Formulation	262
10.5	Basic Solution Approaches	263
10.6	Concept Aggregation — Similarity Based Approach (SBA)	264
10.6.1	SBA — Procedure	264
10.6.2	Meta-Data Vectors	265

10.6.3 SBA — Detailed Illustration	266
10.7 Concept Aggregation — Error Based Approach (eBA) . . .	268
10.7.1 The Confusion Matrix	269
10.7.2 EBA Method — Optimal Solution	270
10.7.3 EBA Method — Heuristic Greedy Solution	273
10.7.3.1 Detailed Procedure	274
10.7.3.2 Illustration	275
10.8 Experimental Study	276
10.8.1 Solving the Optimization Problem of the EBA Opti- mal Method	276
10.8.2 Experiment Results	276
10.9 Concept De-Aggregation	279
10.9.1 Concept De-Aggregation — Outline	280
10.9.2 Models Generation Procedure	280
10.9.3 Termination Condition	281
10.9.4 Finding the Optimal Partition	281
10.10 Conclusions	281
11. A Meta-Classification for Decomposition Methodology	285
11.1 Meta Classifier Schema	285
11.2 Dataset Characterizer	287
11.3 New Entropy Measure for Characterizing Mixed Datasets	290
11.4 Dataset Manipulator	296
11.5 Decomposer Evaluator	297
11.6 The Meta-Decomposer	298
11.7 Evaluating the Meta-Decomposer	298
<i>Bibliography</i>	301
<i>Index</i>	321