

Contents

<i>Preface</i>	vii
1. Introduction to Decision Trees	1
1.1 Data Mining and Knowledge Discovery	1
1.2 Taxonomy of Data Mining Methods	3
1.3 Supervised Methods	4
1.3.1 Overview	4
1.4 Classification Trees	5
1.5 Characteristics of Classification Trees	8
1.5.1 Tree Size	9
1.5.2 The hierarchical nature of decision trees	9
1.6 Relation to Rule Induction	11
2. Growing Decision Trees	13
2.0.1 Training Set	13
2.0.2 Definition of the Classification Problem	14
2.0.3 Induction Algorithms	16
2.0.4 Probability Estimation in Decision Trees	16
2.0.4.1 Laplace Correction	17
2.0.4.2 No Match	18
2.1 Algorithmic Framework for Decision Trees	18
2.2 Stopping Criteria	19
3. Evaluation of Classification Trees	21
3.1 Overview	21
3.2 Generalization Error	21

3.2.1	Theoretical Estimation of Generalization Error . . .	22
3.2.2	Empirical Estimation of Generalization Error . . .	23
3.2.3	Alternatives to the Accuracy Measure	24
3.2.4	The F-Measure	25
3.2.5	Confusion Matrix	27
3.2.6	Classifier Evaluation under Limited Resources . . .	28
3.2.6.1	ROC Curves	30
3.2.6.2	Hit Rate Curve	30
3.2.6.3	Qrecall (Quota Recall)	32
3.2.6.4	Lift Curve	32
3.2.6.5	Pearson Correlation Coefficient	32
3.2.6.6	Area Under Curve (AUC)	34
3.2.6.7	Average Hit Rate	35
3.2.6.8	Average Qrecall	35
3.2.6.9	Potential Extract Measure (PEM)	36
3.2.7	Which Decision Tree Classifier is Better?	40
3.2.7.1	McNemar's Test	40
3.2.7.2	A Test for the Difference of Two Proportions	41
3.2.7.3	The Resampled Paired t Test	43
3.2.7.4	The k -fold Cross-validated Paired t Test	43
3.3	Computational Complexity	44
3.4	Comprehensibility	44
3.5	Scalability to Large Datasets	45
3.6	Robustness	47
3.7	Stability	47
3.8	Interestingness Measures	48
3.9	Overfitting and Underfitting	49
3.10	"No Free Lunch" Theorem	50
4.	Splitting Criteria	53
4.1	Univariate Splitting Criteria	53
4.1.1	Overview	53
4.1.2	Impurity based Criteria	53
4.1.3	Information Gain	54
4.1.4	Gini Index	55
4.1.5	Likelihood Ratio Chi-squared Statistics	55
4.1.6	DKM Criterion	55
4.1.7	Normalized Impurity-based Criteria	56

4.1.8	Gain Ratio	56
4.1.9	Distance Measure	56
4.1.10	Binary Criteria	57
4.1.11	Twoing Criterion	57
4.1.12	Orthogonal Criterion	58
4.1.13	Kolmogorov–Smirnov Criterion	58
4.1.14	AUC Splitting Criteria	58
4.1.15	Other Univariate Splitting Criteria	59
4.1.16	Comparison of Univariate Splitting Criteria	59
4.2	Handling Missing Values	59
5.	Pruning Trees	63
5.1	Stopping Criteria	63
5.2	Heuristic Pruning	63
5.2.1	Overview	63
5.2.2	Cost Complexity Pruning	64
5.2.3	Reduced Error Pruning	65
5.2.4	Minimum Error Pruning (MEP)	65
5.2.5	Pessimistic Pruning	65
5.2.6	Error-Based Pruning (EBP)	66
5.2.7	Minimum Description Length (MDL) Pruning	67
5.2.8	Other Pruning Methods	67
5.2.9	Comparison of Pruning Methods	68
5.3	Optimal Pruning	68
6.	Advanced Decision Trees	71
6.1	Survey of Common Algorithms for Decision Tree Induction	71
6.1.1	ID3	71
6.1.2	C4.5	71
6.1.3	CART	71
6.1.4	CHAID	72
6.1.5	QUEST	73
6.1.6	Reference to Other Algorithms	73
6.1.7	Advantages and Disadvantages of Decision Trees	73
6.1.8	Oblivious Decision Trees	76
6.1.9	Decision Trees Inducers for Large Datasets	78
6.1.10	Online Adaptive Decision Trees	79
6.1.11	Lazy Tree	79

6.1.12	Option Tree	80
6.2	Lookahead	82
6.3	Oblique Decision Trees	83
7.	Decision Forests	87
7.1	Overview	87
7.2	Introduction	87
7.3	Combination Methods	90
7.3.1	Weighting Methods	90
7.3.1.1	Majority Voting	90
7.3.1.2	Performance Weighting	91
7.3.1.3	Distribution Summation	91
7.3.1.4	Bayesian Combination	91
7.3.1.5	Dempster–Shafer	92
7.3.1.6	Vogging	92
7.3.1.7	Naïve Bayes	93
7.3.1.8	Entropy Weighting	93
7.3.1.9	Density-based Weighting	93
7.3.1.10	DEA Weighting Method	93
7.3.1.11	Logarithmic Opinion Pool	94
7.3.1.12	Gating Network	94
7.3.1.13	Order Statistics	95
7.3.2	Meta-combination Methods	95
7.3.2.1	Stacking	95
7.3.2.2	Arbiter Trees	97
7.3.2.3	Combiner Trees	99
7.3.2.4	Grading	100
7.4	Classifier Dependency	101
7.4.1	Dependent Methods	101
7.4.1.1	Model-guided Instance Selection	101
7.4.1.2	Incremental Batch Learning	105
7.4.2	Independent Methods	105
7.4.2.1	Bagging	105
7.4.2.2	Wagging	107
7.4.2.3	Random Forest	108
7.4.2.4	Cross-validated Committees	109
7.5	Ensemble Diversity	109
7.5.1	Manipulating the Inducer	110
7.5.1.1	Manipulation of the Inducer’s Parameters	111

7.5.1.2	Starting Point in Hypothesis Space	111
7.5.1.3	Hypothesis Space Traversal	111
7.5.2	Manipulating the Training Samples	112
7.5.2.1	Resampling	112
7.5.2.2	Creation	113
7.5.2.3	Partitioning	113
7.5.3	Manipulating the Target Attribute Representation .	114
7.5.4	Partitioning the Search Space	115
7.5.4.1	Divide and Conquer	116
7.5.4.2	Feature Subset-based Ensemble Methods .	117
7.5.5	Multi-Inducers	121
7.5.6	Measuring the Diversity	122
7.6	Ensemble Size	124
7.6.1	Selecting the Ensemble Size	124
7.6.2	Pre Selection of the Ensemble Size	124
7.6.3	Selection of the Ensemble Size while Training . . .	125
7.6.4	Pruning — Post Selection of the Ensemble Size . .	125
7.6.4.1	Pre-combining Pruning	126
7.6.4.2	Post-combining Pruning	126
7.7	Cross-Inducer	127
7.8	Multistrategy Ensemble Learning	127
7.9	Which Ensemble Method Should be Used?	128
7.10	Open Source for Decision Trees Forests	128
8.	Incremental Learning of Decision Trees	131
8.1	Overview	131
8.2	The Motives for Incremental Learning	131
8.3	The Inefficiency Challenge	132
8.4	The Concept Drift Challenge	133
9.	Feature Selection	137
9.1	Overview	137
9.2	The “Curse of Dimensionality”	137
9.3	Techniques for Feature Selection	140
9.3.1	Feature Filters	141
9.3.1.1	FOCUS	141
9.3.1.2	LVF	141

9.3.1.3	Using One Learning Algorithm as a Filter for Another	141
9.3.1.4	An Information Theoretic Feature Filter	142
9.3.1.5	An Instance Based Approach to Feature Selection – RELIEF	142
9.3.1.6	Simba and G-flip	142
9.3.1.7	Contextual Merit Algorithm	143
9.3.2	Using Traditional Statistics for Filtering	143
9.3.2.1	Mallows Cp	143
9.3.2.2	AIC, BIC and F-ratio	144
9.3.2.3	Principal Component Analysis (PCA)	144
9.3.2.4	Factor Analysis (FA)	145
9.3.2.5	Projection Pursuit	145
9.3.3	Wrappers	145
9.3.3.1	Wrappers for Decision Tree Learners	145
9.4	Feature Selection as a Means of Creating Ensembles	146
9.5	Ensemble Methodology as a Means for Improving Feature Selection	147
9.5.1	Independent Algorithmic Framework	149
9.5.2	Combining Procedure	150
9.5.2.1	Simple Weighted Voting	151
9.5.2.2	Naïve Bayes Weighting using Artificial Contrasts	152
9.5.3	Feature Ensemble Generator	154
9.5.3.1	Multiple Feature Selectors	154
9.5.3.2	Bagging	156
9.6	Using Decision Trees for Feature Selection	156
9.7	Limitation of Feature Selection Methods	157
10.	Fuzzy Decision Trees	159
10.1	Overview	159
10.2	Membership Function	160
10.3	Fuzzy Classification Problems	161
10.4	Fuzzy Set Operations	163
10.5	Fuzzy Classification Rules	164
10.6	Creating Fuzzy Decision Tree	164
10.6.1	Fuzzifying Numeric Attributes	165
10.6.2	Inducing of Fuzzy Decision Tree	166
10.7	Simplifying the Decision Tree	169

10.8	Classification of New Instances	169
10.9	Other Fuzzy Decision Tree Inducers	169
11.	Hybridization of Decision Trees with other Techniques	171
11.1	Introduction	171
11.2	A Decision Tree Framework for Instance-Space Decomposition	171
11.2.1	Stopping Rules	174
11.2.2	Splitting Rules	175
11.2.3	Split Validation Examinations	175
11.3	The CPOM Algorithm	176
11.3.1	CPOM Outline	176
11.3.2	The Grouped Gain Ratio Splitting Rule	177
11.4	Induction of Decision Trees by an Evolutionary Algorithm	179
12.	Sequence Classification Using Decision Trees	187
12.1	Introduction	187
12.2	Sequence Representation	187
12.3	Pattern Discovery	188
12.4	Pattern Selection	190
12.4.1	Heuristics for Pattern Selection	190
12.4.2	Correlation based Feature Selection	191
12.5	Classifier Training	191
12.5.1	Adjustment of Decision Trees	192
12.5.2	Cascading Decision Trees	192
12.6	Application of CREDIT in Improving of Information Retrieval of Medical Narrative Reports	193
12.6.1	Related Works	195
12.6.1.1	Text Classification	195
12.6.1.2	Part-of-speech Tagging	198
12.6.1.3	Frameworks for Information Extraction	198
12.6.1.4	Frameworks for Labeling Sequential Data	199
12.6.1.5	Identifying Negative Context in Non-domain Specific Text (General NLP)	199
12.6.1.6	Identifying Negative Context in Medical Narratives	200
12.6.1.7	Works Based on Knowledge Engineering	200
12.6.1.8	Works based on Machine Learning	201

12.6.2	Using CREDIT for Solving the Negation Problem	201
12.6.2.1	The Process Overview	201
12.6.2.2	Step 1: Corpus Preparation	201
12.6.2.3	Step 1.1: Tagging	202
12.6.2.4	Step 1.2: Sentence Boundaries	202
12.6.2.5	Step 1.3: Manual Labeling	203
12.6.2.6	Step 2: Patterns Creation	203
12.6.2.7	Step 3: Patterns Selection	206
12.6.2.8	Step 4: Classifier Training	208
12.6.2.9	Cascade of Three Classifiers	209
	<i>Bibliography</i>	215
	<i>Index</i>	243