
Contents

1	Introduction	1
1.1	Types of Uncertainty	1
1.2	Uncertainty Modeling and Data Mining	4
1.3	Related Works	6
	References	9
2	Induction and Learning	13
2.1	Introduction	13
2.2	Machine Learning	14
2.2.1	Searching in Hypothesis Space	16
2.2.2	Supervised Learning	18
2.2.3	Unsupervised Learning	20
2.2.4	Instance-Based Learning	22
2.3	Data Mining and Algorithms	23
2.3.1	Why Do We Need Data Mining?	24
2.3.2	How Do We do Data Mining?	24
2.3.3	Artificial Neural Networks	25
2.3.4	Support Vector Machines	27
2.4	Measurement of Classifiers	29
2.4.1	ROC Analysis for Classification	30
2.4.2	Area Under the ROC Curve	31
2.5	Summary	34
	References	34
3	Label Semantics Theory	39
3.1	Uncertainty Modeling with Labels	39
3.1.1	Fuzzy Logic	39
3.1.2	Computing with Words	41
3.1.3	Mass Assignment Theory	42
3.2	Label Semantics	44
3.2.1	Epistemic View of Label Semantics	45

3.2.2	Random Set Framework	46
3.2.3	Appropriateness Degrees	50
3.2.4	Assumptions for Data Analysis	51
3.2.5	Linguistic Translation	54
3.3	Fuzzy Discretization	57
3.3.1	Percentile-Based Discretization	58
3.3.2	Entropy-Based Discretization	58
3.4	Reasoning with Fuzzy Labels	61
3.4.1	Conditional Distribution Given Mass Assignments	61
3.4.2	Logical Expressions of Fuzzy Labels	62
3.4.3	Linguistic Interpretation of Appropriate Labels	65
3.4.4	Evidence Theory and Mass Assignment	66
3.5	Label Relations	69
3.6	Summary	73
	References	74
4	Linguistic Decision Trees for Classification	77
4.1	Introduction	77
4.2	Tree Induction	77
4.2.1	Entropy	79
4.2.2	Soft Decision Trees	82
4.3	Linguistic Decision for Classification	82
4.3.1	Branch Probability	85
4.3.2	Classification by LDT	88
4.3.3	Linguistic ID3 Algorithm	90
4.4	Experimental Studies	92
4.4.1	Influence of the Threshold	93
4.4.2	Overlapping Between Fuzzy Labels	95
4.5	Comparison Studies	98
4.6	Merging of Branches	102
4.6.1	Forward Merging Algorithm	103
4.6.2	Dual-Branch LDTs	105
4.6.3	Experimental Studies for Forward Merging	105
4.6.4	ROC Analysis for Forward Merging	109
4.7	Linguistic Reasoning	111
4.7.1	Linguistic Interpretation of an LDT	111
4.7.2	Linguistic Constraints	113
4.7.3	Classification of Fuzzy Data	115
4.8	Summary	117
	References	118

5	Linguistic Decision Trees for Prediction	121
5.1	Prediction Trees	121
5.2	Linguistic Prediction Trees	122
5.2.1	Branch Evaluation	123
5.2.2	Defuzzification	126
5.2.3	Linguistic ID3 Algorithm for Prediction	128
5.2.4	Forward Branch Merging for Prediction	128
5.3	Experimental Studies	130
5.3.1	3D Surface Regression	131
5.3.2	Abalone and Boston Housing Problem	134
5.3.3	Prediction of Sunspots	135
5.3.4	Flood Forecasting	137
5.4	Query Evaluation	143
5.4.1	Single Queries	143
5.4.2	Compound Queries	144
5.5	ROC Analysis for Prediction	145
5.5.1	Predictors and Probabilistic Classifiers	145
5.5.2	AUC Value for Prediction	149
5.6	Summary	152
	References	152
6	Bayesian Methods Based on Label Semantics	155
6.1	Introduction	155
6.2	Naive Bayes	156
6.2.1	Bayes Theorem	157
6.2.2	Fuzzy Naive Bayes	158
6.3	Fuzzy Semi-Naive Bayes	159
6.4	Online Fuzzy Bayesian Prediction	161
6.4.1	Bayesian Methods	161
6.4.2	Online Learning	164
6.5	Bayesian Estimation Trees	165
6.5.1	Bayesian Estimation Given an LDT	165
6.5.2	Bayesian Estimation from a Set of Trees	167
6.6	Experimental Studies	168
6.7	Summary	169
	References	171
7	Unsupervised Learning with Label Semantics	177
7.1	Introduction	177
7.2	Non-Parametric Density Estimation	178
7.3	Clustering	180
7.3.1	Logical Distance	181
7.3.2	Clustering of Mixed Objects	185
7.4	Experimental Studies	187
7.4.1	Logical Distance Example	187

7.4.2	Images and Labels Clustering	190
7.5	Summary	191
	References	192
8	Linguistic FOIL and Multiple Attribute Hierarchy for Decision Making	193
8.1	Introduction	193
8.2	Rule Induction	193
8.3	Multi-Dimensional Label Semantics	196
8.4	Linguistic FOIL	199
8.4.1	Information Heuristics for LFOIL	199
8.4.2	Linguistic Rule Generation	200
8.4.3	Class Probabilities Given a Rule Base	202
8.5	Experimental Studies	203
8.6	Multiple Attribute Decision Making	206
8.6.1	Linguistic Attribute Hierarchies	206
8.6.2	Information Propagation Using LDT	209
8.7	Summary	213
	References	213
9	A Prototype Theory Interpretation of Label Semantics	215
9.1	Introduction	215
9.2	Prototype Semantics for Vague Concepts	217
9.2.1	Uncertainty Measures about the Similarity Neighborhoods Determined by Vague Concepts	217
9.2.2	Relating Prototype Theory and Label Semantics	220
9.2.3	Gaussian-Type Density Function	223
9.3	Vague Information Coarsening in Theory of Prototypes	227
9.4	Linguistic Inference Systems	229
9.5	Summary	231
	References	232
10	Prototype Theory for Learning	235
10.1	Introduction	235
10.1.1	General Rule Induction Process	235
10.1.2	A Clustering Based Rule Coarsening	236
10.2	Linguistic Modeling of Time Series Predictions	238
10.2.1	Mackey-Glass Time Series Prediction	239
10.2.2	Prediction of Sunspots	244
10.3	Summary	250
	References	252

11	Prototype-Based Rule Systems	253
11.1	Introduction	253
11.2	Prototype-Based IF-THEN Rules	254
11.3	Rule Induction Based on Data Clustering and Least-Square Regression	257
11.4	Rule Learning Using a Conjugate Gradient Algorithm	260
11.5	Applications in Prediction Problems	262
11.5.1	Surface Predication	262
11.5.2	Mackey-Glass Time Series Prediction	265
11.5.3	Prediction of Sunspots	269
11.6	Summary	274
	References	274
12	Information Cells and Information Cell Mixture Models	277
12.1	Introduction	277
12.2	Information Cell for Cognitive Representation of Vague Concept Semantics	277
12.3	Information Cell Mixture Model (ICMM) for Semantic Representation of Complex Concept	280
12.4	Learning Information Cell Mixture Model from Data Set	281
12.4.1	Objective Function Based on Positive Density Function	282
12.4.2	Updating Probability Distribution of Information Cells	282
12.4.3	Updating Density Functions of Information Cells	283
12.4.4	Information Cell Updating Algorithm	284
12.4.5	Learning Component Number of ICMM	285
12.5	Experimental Study	286
12.6	Summary	290
	References	290