
List of Figures

1.1	The Key Steps of the Data Mining and Knowledge Discovery Process.	8
1.2	Data Defined in Terms of a Single Attribute.	9
1.3	Data Defined in Terms of Two Attributes.	10
1.4	A Random Sample of Observations Classified in Two Classes.	14
1.5	A Simple Sample of Observations Classified in Two Categories.	15
1.6	A Single Classification Rule as Implied by the Data.	16
1.7	Some Possible Classification Rules for the Data Depicted in Figure 1.4.	17
1.8	The Problem of Selecting a New Observation to Send for Class Determination.	18
2.1	The One Clause At a Time (OCAT) Approach (for the CNF case).	36
2.2	Some Possible Classification Rules for the Data Depicted in Figure 1.4.	37
2.3	The Branch-and-Bound Search for the Illustrative Example.	42
3.1	The Search Tree for the Revised Branch-and-Bound Approach.	60
4.1	The RA1 Heuristic.	78
4.2	The RA2 Heuristic.	82
4.3	Accuracy Rates for Systems S_1 and S_2 When the Heuristic RA1 Is Used on the Wisconsin Breast Cancer Data.	88
4.4	Number of Clauses in Systems S_1 and S_2 When the Heuristic RA1 Is Used on the Wisconsin Breast Cancer Data.	88
4.5	Clauses of Systems S_1 and S_2 When the Entire Wisconsin Breast Cancer Data Are Used.	89
4.6	Accuracy Rates for Systems SA and SB When Heuristic RA2 Is Used on the Wisconsin Breast Cancer Data.	91
4.7	Number of Clauses in Systems SA and SB When Heuristic RA2 Is Used on the Wisconsin Breast Cancer Data.	92

4.8	Using the RA1 Heuristic in Conjunction with the B&B Method.	93
4.9	Percentage of the Time the B&B Was Invoked in the Combined RA1/B&B Method.	96
4.10	Ratio of the Number of Clauses by the RA1/B&B Method and the Number of Clauses by the Stand-Alone B&B Method.	96
4.11	Number of Clauses by the Stand-Alone B&B and the RA1/B&B Method.	97
4.12	Ratio of the Time Used by the Stand-Alone B&B and the Time Used by the RA1/B&B Method.	97
4.13	CPU Times by the Stand-Alone B&B and the RA1/B&B Method.	98
5.1	All Possible Classification Scenarios When the Positive and Negative Models Are Considered.	103
5.2	Flowchart of the Proposed Strategy for Guided Learning.	109
5.3a	Results When “Hidden Logic” Is System 8A.	116
5.3b	Results When “Hidden Logic” Is System 16A.	116
5.3c	Results When “Hidden Logic” Is System 32C.	117
5.3d	Results When “Hidden Logic” Is System 32D.	117
5.4	Comparisons between systems S_{RANDOM} , S_{GUIDED} , and $S_{\text{R-GUIDED}}$ when new examples are considered (system S_{HIDDEN} is $(\bar{A}_1 \vee \bar{A}_4 \vee A_6) \wedge (\bar{A}_2 \vee A_8) \wedge (A_2)$).	118
5.5a	Results When the Breast Cancer Data Are Used. The Focus Is on the Number of Clauses.	121
5.5b	Results When the Breast Cancer Data Are Used. The Focus Is on the Accuracy Rates.	122
6.1	A Sample Training Set of Six Positive Examples and a Set of Four Negative Examples and a Boolean Function Implied by These Data.	127
6.2	Proposed Strategy for Repairing a Boolean Function which Incorrectly Rejects a Positive Example (for the DNF case).	132
6.3	Repair of a Boolean Function that Erroneously Accepts a Negative Example (for the DNF case).	133
6.4	Accuracy Results for the Class-Pair (DOE vs. ZIPFF).	137
6.5	Accuracy Results for the Class-Pair (AP vs. DOE).	137
6.6	Accuracy Results for the Class-Pair (WSJ vs. ZIPFF).	138
6.7	Number of Clauses for the Class-Pair (DOE vs. ZIPFF).	139
6.8	Number of Clauses for the Class-Pair (AP vs. DOE).	140
6.9	Number of Clauses for the Class-Pair (WSJ vs. ZIPFF).	140
6.10	Required CPU Time for the Class-Pair (DOE vs. ZIPFF).	142
6.11	Required CPU Time for the Class-Pair (AP vs. DOE).	142
6.12	Required CPU Time for the Class-Pair (WSJ vs. ZIPFF).	143
8.1	The Rejectability Graph of E^+ and E^-	153
8.2	The Rejectability Graph for the Second Illustrative Example.	154
8.3	The Rejectability Graph for E^+ and E^-	159

9.1	Comparison of the Actual and Computed Borders Between Diagnostic Classes (<i>a Conceptual Representation</i>).	180
9.2	Relations Between Biopsy Class Size and Sample.	182
9.3	Relations Between Cancer Class Size and Sample.	183
10.1	Hierarchical Decomposition of the Breast Cancer Diagnosis Attributes.	197
10.2	The Poset Formed by $\{0, 1\}^4$ and the Relation \preceq	198
10.3	Visualization of a Sample Monotone Boolean Function and Its Values in $\{0, 1\}^4$ ($f(x) = (x_1 \vee x_2) \wedge (x_1 \vee x_3)$).	200
10.4	A Visualization of the Main Idea Behind a Pair of Nested Monotone Boolean Functions.	202
10.5	The Average Query Complexities for Problem 1.	216
10.6	The Average Query Complexities for Problem 2.	217
10.7	Increase in Query Complexities Due to Restricted Access to the Oracles.	218
10.8	Reduction in Query Complexity Due to the Nestedness Assumption.	218
10.9	Average Case Behavior of Various Selection Criteria for Problem 3.	221
10.10	The Restricted and Regular Maximum Likelihood Ratios Simulated with Expected $q = 0.2$ and $n = 3$	222
11.1	A Visualization of a Decomposition of a General Function into General Increasing and Decreasing Functions.	230
11.2	The Data Points in Terms of Attributes X_2 and X_3 Only.	237
11.3	Monotone Discrimination of the Positive (Acceptable) and Negative (Unacceptable) Design Classes.	239
12.1	The RA1 Heuristic for the CNF Case (see also Chapter 4).	245
12.2	The Proposed Altered Randomized Algorithm 1 (ARA1) for the Mining of Association Rules (for the CNF Case).	248
12.3	Histogram of the Results When the Apriori Approach Was Used on Database #2.	250
12.4	Histogram of the Results When the ARA1 Approach Was Used on Database #2.	250
12.5	Histogram of the Results When the MineSet Software Was Used on Database #3.	252
12.6	Histogram of the Results When the ARA1 Approach Was Used on Database #3.	252
12.7	Histogram of the Results When the MineSet Software Was Used on Database #4.	253
12.8	Histogram of the Results When the ARA1 Approach Was Used on Database #4.	253
12.9	Histogram of the Results When the MineSet Software Was Used on Database #5.	254

12.10	Histogram of the Results When the ARA1 Approach Was Used on Database #5.....	254
13.1	A Sample of Four Positive and Six Negative Examples.....	261
13.2	The Training Example Sets in Reverse Roles.	261
13.3	The Vector Space Model (VSM) Approach.....	263
13.4	Comparison of the Classification Decisions Under the VSM and the OCAT/RA1 Approaches.	270
13.5	Results When the GUIDED and RANDOM Approaches Were Used on the (DOE vs. ZIPFF) Class-Pair.	273
13.6	Results When the GUIDED and RANDOM Approaches Were Used on the (AP vs. DOE) Class-Pair.	274
13.7	Results When the GUIDED and RANDOM Approaches Were Used on the (WSJ vs. ZIPFF) Class-Pair.....	274
15.1	A Diagnostic Rule (Rule #2) Inferred from the Breast Cancer Data. .	295
16.1	A Typical Triangular Fuzzy Number.	301
16.2	A Typical Trapezoid Fuzzy Number.	301
16.3	Membership Functions Related to the Number of Undulations.	302
16.4a	A Diagrammatic Representation of a Mass with Undulations.....	303
16.4b	Membership Functions Related to the Length of Undulations.	303
16.5	Fuzzy Logic Structures for a <i>Lobular</i> Mass.	304
16.6	Fuzzy Logic Structures for a <i>Microlobulated</i> Mass.	304
16.7	Structural Descriptions for a Fuzzy Lobular and Microlobulated Mass.	305
16.8	Fuzzy Logic Structures for a Mass with Less Than Three Undulations.....	305
16.9	Diagrammatic Representation of Masses with (a) Deep and (b) Shallow Undulations.	306
17.1	The “Data Mining and Knowledge Discovery Equation for the Future.”.....	314