

# Contents

Preface .....	v
<b>1. Introduction .....</b>	<b>1</b>
<b>2. Search Space Partition-Based Rule Mining .....</b>	<b>5</b>
2.1 Problem Statement .....	5
2.1.1 Canonical Attribute Sequences ( <i>cas</i> ) .....	5
2.1.2 Database .....	6
2.1.3 Support .....	7
2.1.4 Association Rule .....	7
2.1.5 Problem Statement .....	9
2.2 Search Space .....	9
2.3 Splitting Procedure .....	11
2.4 Enumerating $\sigma$ -Frequent Attribute Sets ( <i>cas</i> ) .....	13
2.5 Sequential Enumeration Procedure .....	17
2.6 Parallel Enumeration Procedure .....	19
2.6.1 Initial Load Balancing .....	19
2.6.2 Computing the Starting Sets .....	20
2.6.3 Enumeration Procedure .....	23
2.6.4 Dynamic Load Balancing .....	23
2.7 Generating the Association Rules .....	24
2.7.1 Sequential Generation .....	26
2.7.2 Parallel Generation .....	27
<b>3. Apriori and Other Algorithms .....</b>	<b>33</b>
3.1 Early Algorithms .....	33
3.1.1 AIS .....	33
3.1.2 SETM .....	33
3.2 The Apriori Algorithms .....	34
3.2.1 Apriori .....	34
3.2.2 AprioriTid .....	38
3.3 Direct Hashing and Pruning .....	41
3.3.1 Filtering Candidates .....	41

---

3.3.2	<i>Database Trimming</i> .....	42
3.3.3	<i>The DHP Algorithm</i> .....	43
3.4	<i>Dynamic Set Counting</i> .....	46
<b>4.</b>	<b>Mining for Rules over Attribute Taxonomies</b> .....	<b>49</b>
4.1	<i>Association Rules over Taxonomies</i> .....	50
4.2	<i>Problem Statement and Algorithms</i> .....	54
4.3	<i>Pruning Uninteresting Rules</i> .....	59
4.3.1	<i>Measure of Interest</i> .....	59
4.3.2	<i>Rule Pruning Algorithm</i> .....	61
4.3.3	<i>Attribute Presence-Based Pruning</i> .....	64
<b>5.</b>	<b>Constraint-Based Rule Mining</b> .....	<b>67</b>
5.1	<i>Boolean Constraints</i> .....	67
5.1.1	<i>Syntax</i> .....	67
5.1.2	<i>Semantics</i> .....	68
5.1.3	<i>Propagation of Boolean Constraints</i> .....	70
5.2	<i>Prime Implicants</i> .....	70
5.3	<i>Problem Statement and Algorithms</i> .....	73
<b>6.</b>	<b>Data Partition-Based Rule Mining</b> .....	<b>79</b>
6.1	<i>Data Partitioning</i> .....	79
6.1.1	<i>Building a Probabilistic Model</i> .....	80
6.1.2	<i>Bounding Large Deviations for One cas (Chernoff bounds)</i> .....	81
6.1.3	<i>Bounding Large Deviations for Sets of cass</i> .....	82
6.2	<i>cas Enumeration with Partitioned Data</i> .....	89
6.2.1	<i>Data Partitioning</i> .....	89
6.2.2	<i>Local <math>\sigma</math>-Frequent cas Generation</i> .....	89
6.2.3	<i>Global <math>\sigma</math>-Frequent cas Generation</i> .....	90
<b>7.</b>	<b>Mining for Rules with Categorical and Metric Attributes</b> .....	<b>93</b>
7.1	<i>Interval Systems and Quantitative Rules</i> .....	95
7.2	<i>k-Partial Completeness</i> .....	99
7.3	<i>Pruning Uninteresting Rules</i> .....	102
7.3.1	<i>Measure of Interest</i> .....	103
7.3.2	<i>Attribute Presence-Based Pruning</i> .....	108
7.4	<i>Enumeration Algorithms</i> .....	109
<b>8.</b>	<b>Optimizing Rules with Quantitative Attributes</b> .....	<b>111</b>
8.1	<i>Solving 1-1-Type Rule Optimization Problems</i> .....	113
8.1.1	<i>Problem Statement</i> .....	113
8.1.2	<i>MCS Problem</i> .....	114
8.1.3	<i>MS\C Problem</i> .....	121
8.1.4	<i>MG Problem</i> .....	125

8.2	Solving d-1-Type Rule Optimization Problems.....	125
8.3	Solving 1-q-Type Rule Optimization Problems.....	126
8.3.1	<i>Problem Statement</i> .....	126
8.3.2	<i>MS\C Problem</i> .....	128
8.3.3	<i>MG Problem</i> .....	138
8.4	Solving d-q-Type Rule Optimization Problems.....	144
8.4.1	<i>Problem Statement</i> .....	144
8.4.2	<i>Basic Enumeration</i> .....	146
8.4.3	<i>Enumeration with Pruning</i> .....	147
8.4.4	<i>Pruning the Instantiation Set</i> .....	150
<b>9.</b>	<b>Beyond Support-Confidence Framework.....</b>	<b>151</b>
9.1	A Criticism of the Support-Confidence Framework.....	151
9.2	Conviction.....	153
9.3	Pruning Conviction-Based Rules.....	157
9.3.1	<i>Analyzing Conviction</i> .....	157
9.3.2	<i>Transitivity-Based Pruning</i> .....	158
9.3.3	<i>Improvement-Based Pruning</i> .....	158
9.4	One-Step Association Rule Mining.....	159
9.4.1	<i>Building a Procedure for One-Step Mining</i> .....	160
9.4.2	<i>Building a Procedure for Improvement-Based Pruning</i> .....	164
9.5	Correlated Attribute-Set Mining.....	167
9.5.1	<i>Collective Strength</i> .....	167
9.5.2	<i>Correlated Attribute-Set Enumeration</i> .....	172
9.6	Refining Conviction: Association Rule Intensity.....	178
9.6.1	<i>Measure Construction</i> .....	178
9.6.2	<i>Properties</i> .....	181
9.6.3	<i>Relating <math>\alpha</math>-int(<math>s \Rightarrow u</math>) to conv(<math>s \Rightarrow u</math>)</i> .....	181
9.6.4	<i>Mining with the Intensity Measure</i> .....	182
9.6.5	<i><math>\alpha</math>-Intensity Versus Intensity as Defined in [G96]</i> .....	183
<b>10.</b>	<b>Search Space Partition-Based Sequential Pattern Mining....</b>	<b>185</b>
10.1	Problem Statement.....	185
10.1.1	<i>Sequences of cass</i> .....	185
10.1.2	<i>Database</i> .....	186
10.1.3	<i>Support</i> .....	187
10.1.4	<i>Problem Statement</i> .....	189
10.2	Search Space.....	189
10.3	Splitting the Search Space.....	190
10.4	Splitting Procedure.....	195
10.5	Sequence Enumeration.....	200
10.5.1	<i>Extending the Support Set Notion</i> .....	201
10.5.2	<i>Join Operations</i> .....	202
10.5.3	<i>Sequential Enumeration Procedure</i> .....	208
10.5.4	<i>Parallel Enumeration Procedure</i> .....	215

**Appendix 1. Chernoff Bounds ..... 229**

**Appendix 2. Partitioning in Figure 10.5: Beyond 3rd Power ..... 233**

**Appendix 3. Partitioning in Figure 10.6: Beyond 3rd Power ..... 237**

**References..... 245**

**Index..... 251**