

# Contents

<b>1</b>	<b>An Introduction to Frequent Pattern Mining</b>	<b>1</b>
	Charu C. Aggarwal	
1	Introduction	1
2	Frequent Pattern Mining Algorithms	3
2.1	Frequent Pattern Mining with the Traditional Support Framework	4
2.2	Interesting and Negative Frequent Patterns	6
2.3	Constrained Frequent Pattern Mining	7
2.4	Compressed Representations of Frequent Patterns	7
3	Scalability Issues in Frequent Pattern Mining	8
3.1	Frequent Pattern Mining in Data Streams	8
3.2	Frequent Pattern Mining with Big Data	9
4	Frequent Pattern Mining with Advanced Data Types	9
4.1	Sequential Pattern Mining	10
4.2	Spatiotemporal Pattern Mining	10
4.3	Frequent Patterns in Graphs and Structured Data	11
4.4	Frequent Pattern Mining with Uncertain Data	11
5	Privacy Issues	12
6	Applications of Frequent Pattern Mining	13
6.1	Applications to Major Data Mining Problems	13
6.2	Generic Applications	13
7	Conclusions and Summary	14
	References	14
<b>2</b>	<b>Frequent Pattern Mining Algorithms: A Survey</b>	<b>19</b>
	Charu C. Aggarwal, Mansurul A. Bhuiyan and Mohammad Al Hasan	
1	Introduction	19
1.1	Definitions	22
2	Join-Based Algorithms	23
2.1	Apriori Method	24
2.2	DHP Algorithm	27
2.3	Special Tricks for 2-Itemset Counting	28

2.4	Pruning by Support Lower Bounding . . . . .	28
2.5	Hypercube Decomposition . . . . .	29
3	Tree-Based Algorithms . . . . .	29
3.1	AIS Algorithm . . . . .	31
3.2	TreeProjection Algorithms . . . . .	32
3.3	Vertical Mining Algorithms . . . . .	36
4	Recursive Suffix-Based Growth . . . . .	39
4.1	The FP-Growth Approach . . . . .	41
4.2	Variations . . . . .	45
5	Maximal and Closed Frequent Itemsets . . . . .	47
5.1	Definitions . . . . .	47
5.2	Frequent Maximal Itemset Mining Algorithms . . . . .	48
5.3	Frequent Closed Itemset Mining Algorithms . . . . .	55
6	Other Optimizations and Variations . . . . .	57
6.1	Row Enumeration Methods . . . . .	57
6.2	Other Exploration Strategies . . . . .	58
7	Reducing the Number of Passes . . . . .	58
7.1	Combining Passes . . . . .	58
7.2	Sampling Tricks . . . . .	59
7.3	Online Association Rule Mining . . . . .	60
8	Conclusions and Summary . . . . .	61
	References . . . . .	61
<b>3</b>	<b>Pattern-Growth Methods . . . . .</b>	<b>65</b>
	Jiawei Han and Jian Pei	
1	Introduction . . . . .	66
2	FP-Growth: Pattern Growth for Mining Frequent Itemsets . . . . .	68
3	Pushing More Constraints in Pattern-Growth Mining . . . . .	72
4	PrefixSpan: Mining Sequential Patterns by Pattern Growth . . . . .	74
5	Further Development of Pattern Growth-Based Pattern Mining Methodology . . . . .	77
6	Conclusions . . . . .	78
	References . . . . .	79
<b>4</b>	<b>Mining Long Patterns . . . . .</b>	<b>83</b>
	Feida Zhu	
1	Introduction . . . . .	83
2	Preliminaries . . . . .	84
3	A Pattern Lattice Model . . . . .	86
4	Pattern Enumeration Approach . . . . .	87
4.1	Breadth-First Approach . . . . .	87
4.2	Depth-First Approach . . . . .	88
5	Row Enumeration Approach . . . . .	89
6	Pattern Merge Approach . . . . .	92
6.1	Piece-wise Pattern Merge . . . . .	93

6.2	Fusion-style Pattern Merge .....	98
7	Pattern Traversal Approach .....	101
8	Conclusion .....	102
	References .....	103
<b>5</b>	<b>Interesting Patterns</b> .....	<b>105</b>
	Jilles Vreeken and Nikolaj Tatti	
1	Introduction .....	106
2	Absolute Measures .....	107
2.1	Frequent Itemsets .....	107
2.2	Tiles .....	112
2.3	Low Entropy Sets .....	114
3	Advanced Methods .....	114
4	Static Background Models .....	115
4.1	Independence Model .....	116
4.2	Beyond Independence .....	119
4.3	Maximum Entropy Models .....	120
4.4	Randomization Approaches .....	123
5	Dynamic Background Models .....	124
5.1	The General Idea .....	125
5.2	Maximum Entropy Models .....	125
5.3	Tile-based Techniques .....	126
5.4	Swap Randomization .....	128
6	Pattern Sets .....	128
6.1	Itemsets .....	129
6.2	Tiles .....	130
6.3	Swap Randomization .....	130
7	Conclusions .....	131
	References .....	132
<b>6</b>	<b>Negative Association Rules</b> .....	<b>135</b>
	Luiza Antonie, Jundong Li and Osmar Zaiane	
1	Introduction .....	135
2	Negative Patterns and Negative Association Rules .....	136
3	Current Approaches .....	138
4	Associative Classification and Negative Association Rules .....	143
5	Conclusions .....	143
	References .....	144
<b>7</b>	<b>Constraint-Based Pattern Mining</b> .....	<b>147</b>
	Siegfried Nijssen and Albrecht Zimmermann	
1	Introduction .....	147
2	Problem Definition .....	148
2.1	Constraints .....	149
3	Level-Wise Algorithm .....	152

3.1	Generic Algorithm .....	153
4	Depth-First Algorithm .....	154
4.1	Basic Algorithm .....	154
4.2	Constraint-based Itemset Mining .....	155
4.3	Generic Frameworks .....	158
4.4	Implementation Considerations .....	159
5	Languages .....	159
6	Conclusions .....	162
	References .....	162
<b>8</b>	<b>Mining and Using Sets of Patterns through Compression .....</b>	<b>165</b>
	Matthijs van Leeuwen and Jilles Vreeken	
1	Introduction .....	165
2	Foundations .....	167
2.1	Kolmogorov Complexity .....	168
2.2	MDL .....	169
2.3	MDL in Data Mining .....	171
3	Compression-based Pattern Models .....	171
3.1	Pattern Models for MDL .....	172
3.2	Code Tables .....	173
3.3	Instances of Compression-based Models .....	179
4	Algorithmic Approaches .....	181
4.1	Candidate Set Filtering .....	181
4.2	Direct Mining of Patterns that Compress .....	184
5	MDL for Data Mining .....	185
5.1	Classification .....	186
5.2	A Dissimilarity Measure for Datasets .....	188
5.3	Identifying and Characterizing Components .....	189
5.4	Other Data Mining Tasks .....	191
5.5	The Advantage of Pattern-based Models .....	192
6	Challenges Ahead .....	193
6.1	Toward Mining Structured Data .....	193
6.2	Generalization .....	194
6.3	Task- and/or User-specific Usefulness .....	194
7	Conclusions .....	195
	References .....	196
<b>9</b>	<b>Frequent Pattern Mining in Data Streams .....</b>	<b>199</b>
	Victor E. Lee, Ruoming Jin and Gagan Agrawal	
1	Introduction .....	200
2	Preliminaries .....	201
2.1	Frequent Pattern Mining: Definition .....	201
2.2	Data Windows .....	202
2.3	Frequent Item Mining .....	203
3	Frequent Itemset Mining Algorithms .....	204

3.1	Mining the Full Data Stream .....	206
3.2	Recently Frequent Itemsets .....	209
3.3	Closed and Maximal Itemsets .....	214
3.4	Mining Data Streams with Uncertain Data .....	216
4	Mining Patterns Other than Itemsets .....	216
4.1	Subsequences .....	217
4.2	Subtrees and Semistructured Data .....	218
4.3	Subgraphs .....	219
5	Concluding Remarks .....	219
	References .....	220
<b>10</b>	<b>Big Data Frequent Pattern Mining .....</b>	<b>225</b>
	David C. Anastasiu, Jeremy Iverson, Shaden Smith and George Karypis	
1	Introduction .....	225
2	Frequent Pattern Mining: Overview .....	226
2.1	Preliminaries .....	226
2.2	Basic Mining Methodologies .....	228
3	Paradigms for Big Data Computation .....	232
3.1	Principles of Parallel Algorithms .....	232
3.2	Shared Memory Systems .....	233
3.3	Distributed Memory Systems .....	234
4	Frequent Itemset Mining .....	236
4.1	Memory Scalability .....	236
4.2	Work Partitioning .....	239
4.3	Dynamic Load Balancing .....	241
4.4	Further Considerations .....	242
5	Frequent Sequence Mining .....	242
5.1	Serial Frequent Sequence Mining .....	243
5.2	Parallel Frequent Sequence Mining .....	245
6	Frequent Graph Mining .....	250
6.1	Serial Frequent Graph Mining .....	250
6.2	Parallel Frequent Graph Mining .....	252
7	Conclusion .....	255
	References .....	256
<b>11</b>	<b>Sequential Pattern Mining .....</b>	<b>261</b>
	Wei Shen, Jianyong Wang and Jiawei Han	
1	Introduction .....	261
2	Problem Definition .....	263
3	Apriori-based Approaches .....	264
3.1	Horizontal Data Format Algorithms .....	264
3.2	Vertical Data Format Algorithms .....	268
4	Pattern Growth Algorithms .....	271
4.1	FreeSpan .....	271
4.2	PrefixSpan .....	272

5	Extensions . . . . .	274
5.1	Closed Sequential Pattern Mining . . . . .	274
5.2	Multi-level, Multi-dimensional Sequential Pattern Mining . . .	276
5.3	Incremental Methods . . . . .	277
5.4	Hybrid Methods . . . . .	278
5.5	Approximate Methods . . . . .	279
5.6	Top- $k$ Closed Sequential Pattern Mining . . . . .	279
5.7	Frequent Episode Mining . . . . .	280
6	Conclusions and Summary . . . . .	281
	References . . . . .	281
<b>12</b>	<b>Spatiotemporal Pattern Mining: Algorithms and Applications . . . . .</b>	<b>283</b>
	Zhenhui Li	
1	Introduction . . . . .	283
2	Basic Concept . . . . .	284
2.1	Spatiotemporal Data Collection . . . . .	284
2.2	Data Preprocessing . . . . .	285
2.3	Background Information . . . . .	286
3	Individual Periodic Pattern . . . . .	286
3.1	Automatic Discovery of Periodicity in Movements . . . . .	287
3.2	Frequent Periodic Pattern Mining . . . . .	289
3.3	Using Periodic Pattern for Location Prediction . . . . .	289
4	Pairwise Movement Patterns . . . . .	290
4.1	Similarity Measure . . . . .	290
4.2	Generic Pattern . . . . .	292
4.3	Behavioral Pattern . . . . .	294
4.4	Semantic Patterns . . . . .	296
5	Aggregate Patterns over Multiple Trajectories . . . . .	298
5.1	Frequent Trajectory Pattern Mining . . . . .	298
5.2	Detection of Moving Object Cluster . . . . .	300
5.3	Trajectory Clustering . . . . .	302
6	Summary . . . . .	304
	References . . . . .	304
<b>13</b>	<b>Mining Graph Patterns . . . . .</b>	<b>307</b>
	Hong Cheng, Xifeng Yan and Jiawei Han	
1	Introduction . . . . .	307
2	Frequent Subgraph Mining . . . . .	308
2.1	Problem Definition . . . . .	308
2.2	Apriori-Based Approach . . . . .	309
2.3	Pattern-Growth Approach . . . . .	310
2.4	Closed and Maximal Subgraphs . . . . .	311
2.5	Mining Subgraphs in a Single Graph . . . . .	311
2.6	The Computational Bottleneck . . . . .	313

3	Mining Significant Graph Patterns .....	314
3.1	Problem Definition .....	314
3.2	gboost: A Branch-and-Bound Approach .....	314
3.3	gPLS: A Partial Least Squares Regression Approach .....	317
3.4	LEAP: A Structural Leap Search Approach .....	319
3.5	GraphSig: A Feature Representation Approach .....	323
4	Mining Representative Orthogonal Graphs .....	326
4.1	Problem Definition .....	327
4.2	Randomized Maximal Subgraph Mining .....	327
4.3	Orthogonal Representative Set Generation .....	329
5	Mining Dense Graph Patterns .....	329
5.1	Cliques and Quasi-Cliques .....	330
5.2	K-Core and K-Truss .....	331
5.3	Other Dense Subgraph Patterns .....	332
6	Mining Graph Patterns in Streams .....	332
7	Mining Graph Patterns in Uncertain Graphs .....	334
8	Conclusions .....	336
	References .....	336
<b>14</b>	<b>Uncertain Frequent Pattern Mining .....</b>	<b>339</b>
	Carson Kai-Sang Leung	
1	Introduction .....	339
2	The Probabilistic Model for Mining Expected Support-Based Frequent Patterns from Uncertain Data .....	340
3	Candidate Generate-and-Test Based Uncertain Frequent Pattern Mining .....	343
4	Hyperlinked Structure-Based Uncertain Frequent Pattern Mining ..	344
5	Tree-Based Uncertain Frequent Pattern Mining .....	345
5.1	UF-growth .....	345
5.2	UFP-growth .....	346
5.3	CUF-growth .....	347
5.4	PUF-growth .....	349
6	Constrained Uncertain Frequent Pattern Mining .....	350
7	Uncertain Frequent Pattern Mining from Big Data .....	351
8	Streaming Uncertain Frequent Pattern Mining .....	353
8.1	SUF-growth .....	353
8.2	UF-streaming for the Sliding Window Model .....	354
8.3	TUF-streaming for the Time-Fading Model .....	355
8.4	LUF-streaming for the Landmark Model .....	356
8.5	Hyperlinked Structure-Based Streaming Uncertain Frequent Pattern Mining .....	356
9	Vertical Uncertain Frequent Pattern Mining .....	357
9.1	U-Eclat: An Approximate Algorithm .....	357
9.2	UV-Eclat: An Exact Algorithm .....	357
9.3	U-VIPER: An Exact Algorithm .....	358

10	Discussion on Uncertain Frequent Pattern Mining . . . . .	360
11	Extension: Probabilistic Frequent Pattern Mining . . . . .	361
11.1	Mining Probabilistic Heavy Hitters . . . . .	361
11.2	Mining Probabilistic Frequent Patterns . . . . .	362
12	Conclusions . . . . .	364
	References . . . . .	365
<b>15</b>	<b>Privacy Issues in Association Rule Mining . . . . .</b>	<b>369</b>
	Aris Gkoulalas-Divanis, Jayant Haritsa and Murat Kantarcioglu	
1	Introduction . . . . .	369
2	Input Privacy . . . . .	370
2.1	Problem Framework . . . . .	371
2.2	Evolution of the Literature . . . . .	376
3	Output Privacy . . . . .	379
3.1	Terminology and Preliminaries . . . . .	380
3.2	Taxonomy of ARH Algorithms . . . . .	381
3.3	Heuristic and Exact ARH Algorithms . . . . .	382
3.4	Metrics and Performance Analysis . . . . .	390
4	Cryptographic Methods . . . . .	392
4.1	Horizontally Partitioned Data . . . . .	394
4.2	Vertically Partitioned Data . . . . .	396
5	Conclusions . . . . .	398
	References . . . . .	398
<b>16</b>	<b>Frequent Pattern Mining Algorithms for Data Clustering . . . . .</b>	<b>403</b>
	Arthur Zimek, Ira Assent and Jilles Vreeken	
1	Introduction . . . . .	403
2	Generalizing Pattern Mining for Clustering . . . . .	406
2.1	Generalized Monotonicity . . . . .	407
2.2	Count Indexes . . . . .	410
2.3	Pattern Explosion and Redundancy . . . . .	410
3	Frequent Pattern Mining in Subspace Clustering . . . . .	412
3.1	Subspace Cluster Search . . . . .	412
3.2	Subspace Search . . . . .	414
3.3	Redundancy in Subspace Clustering . . . . .	417
4	Conclusions . . . . .	419
	References . . . . .	419
<b>17</b>	<b>Supervised Pattern Mining and Applications to Classification . . . . .</b>	<b>425</b>
	Albrecht Zimmermann and Siegfried Nijssen	
1	Introduction . . . . .	425
2	Supervised Pattern Mining . . . . .	427
2.1	Explicit Class Labels . . . . .	428
2.2	Classes as Data Subsets . . . . .	428
2.3	Numerical Target Values . . . . .	431

3	Supervised Pattern Set Mining . . . . .	432
3.1	Local Evaluation, Local Modification . . . . .	434
3.2	Global Evaluation, Global Modification . . . . .	435
3.3	Local Evaluation, Global Modification . . . . .	436
3.4	Data Instance-Based Selection . . . . .	437
4	Classifier Construction . . . . .	437
4.1	Direct Classification . . . . .	437
4.2	Indirect Classification . . . . .	438
5	Summary . . . . .	439
	References . . . . .	440
<b>18</b>	<b>Applications of Frequent Pattern Mining . . . . .</b>	<b>443</b>
	Charu C. Aggarwal	
1	Introduction . . . . .	443
2	Frequent Patterns for Customer Analysis . . . . .	445
3	Frequent Patterns for Clustering . . . . .	446
4	Frequent Patterns for Classification . . . . .	447
5	Frequent Patterns for Outlier Analysis . . . . .	449
6	Frequent Patterns for Indexing . . . . .	450
7	Web Mining Applications . . . . .	451
7.1	Web Log Mining . . . . .	451
7.2	Web Linkage Mining . . . . .	452
8	Frequent Patterns for Text Mining . . . . .	452
9	Temporal Applications . . . . .	453
10	Spatial and Spatiotemporal Applications . . . . .	455
11	Software Bug Detection . . . . .	456
12	Chemical and Biological Applications . . . . .	457
12.1	Chemical Applications . . . . .	458
12.2	Biological Applications . . . . .	458
13	Resources for the Practitioner . . . . .	460
14	Conclusions and Summary . . . . .	461
	References . . . . .	461
	<b>Index . . . . .</b>	<b>469</b>

Frequent Pattern Mining

Aggarwal, C.C.; Han, J. (Eds.)

2014, XIX, 471 p. 83 illus., 19 illus. in color., Hardcover

ISBN: 978-3-319-07820-5