

Contents

1	Introduction	1
1.1	Background	1
1.2	The Proposed Book	2
1.2.1	Mining Market Basket Database	2
1.2.2	Mining Time-Stamped Databases	3
1.2.3	Mining Multiple Related Databases	4
1.3	Experimental Settings	6
1.4	Future Directions	6
	References	7
2	Synthesizing Conditional Patterns in a Database	9
2.1	Introduction	9
2.2	Conditional Pattern	11
2.3	Properties of Conditional Patterns	14
2.4	Mining Conditional Patterns	17
2.4.1	Algorithm Design	19
2.5	Experiments	21
2.5.1	An Application	26
2.6	Related Work	28
2.7	Conclusion and Future Work	29
	References	29
3	Synthesizing Arbitrary Boolean Expressions Induced by Frequent Itemsets	31
3.1	Introduction	31
3.2	Related Results	35
3.3	Synthesizing Generators	37
3.3.1	Algorithm Design	38
3.3.2	Synthesizing First k Boolean Expressions Induced by Top p Frequent Itemsets	40

3.4	Experiments	41
3.5	Related Work	45
3.6	Conclusion.	46
	References.	46
4	Measuring Association Among Items in a Database	49
4.1	Introduction	49
4.2	Related Work	51
4.3	New Measures of Association	52
4.4	Properties of A_1 and A_2	54
4.4.1	Capturing Association	63
4.5	An Application: Clustering Frequent Items in a Database.	64
4.6	Experimental Results.	64
4.6.1	Top Associative Itemsets Among Frequent Itemsets.	65
4.6.2	Finding the Difference in Similarity	65
4.6.3	Execution Time for Measuring Association	68
4.7	Conclusion.	71
	References.	71
5	Mining Association Rules Induced by Item and Quantity Purchased	73
5.1	Introduction	73
5.2	Association Rules in a TIMT Type Database	74
5.3	Frameworks for Mining Association Rules.	75
5.3.1	Framework for Mining Association Rules Under Category I	76
5.3.2	Framework for Mining Association Rules Under Category II	76
5.3.3	Framework for Mining Association Rules Under Category III.	77
5.3.4	Dealing with Items Measured in Continuous Scale	77
5.4	Properties of Different Interestingness Measures	78
5.5	Mining Association Rules	78
5.6	Experiments	79
5.7	Related Work.	81
5.8	Conclusion.	81
	References.	81
6	Mining Patterns in Different Related Databases	83
6.1	Introduction	83
6.2	Multi-database Mining Using Local Pattern Analysis.	84
6.3	Generalized Multi-database Mining Techniques	85

6.3.1	Local Pattern Analysis.	85
6.3.2	Partition Algorithm.	86
6.3.3	IdentifyExPattern Algorithm.	86
6.3.4	RuleSynthesizing Algorithm.	87
6.4	Specialized Multi-database Mining Techniques.	87
6.4.1	Mining Multiple Real Databases.	87
6.4.2	Mining Multiple Databases for the Purpose of Studying a Set of Items.	88
6.4.3	Study of Temporal Patterns in Multiple Databases.	88
6.5	Mining Multiple Databases Using Pipelined Feedback Model (PFM).	89
6.5.1	Algorithm Design.	90
6.6	Error Evaluation.	90
6.7	Experiments.	91
6.8	Conclusions.	95
	References.	95
7	Mining Icebergs in Different Time-Stamped Data Sources.	97
7.1	Introduction.	97
7.2	Related Work.	99
7.3	Notches in Sales Series.	100
7.3.1	Identifying Notches.	102
7.4	Generalized Notch.	103
7.5	Iceberg Notch.	104
7.6	Sales Series.	104
7.6.1	Another View at Sales Series.	104
7.6.2	Other Applications of Icebergs.	105
7.7	Mining Icebergs in Time-Stamped Databases.	106
7.7.1	Non-incremental Approach.	106
7.7.2	Incremental Approach.	109
7.8	Significant Year.	111
7.9	Experimental Studies.	112
7.10	Conclusions.	124
	References.	125
8	Synthesizing Exceptional Patterns in Different Data Sources.	127
8.1	Introduction.	127
8.2	Exceptional Patterns in Multiple Data Sources.	129
8.3	Problem Statement.	135
8.4	Related Work.	136
8.5	Synthesizing Support of an Itemset.	137
8.6	Synthesizing Type II Global Exceptional Itemsets.	138
8.7	Error Calculation.	142

8.8	Experiments	143
8.8.1	Comparison with the Existing Algorithm	147
8.9	Conclusions	151
	References.	152
9	Clustering Items in Time-Stamped Databases.	155
9.1	Introduction	155
9.2	Related Work	156
9.3	A Model of Mining Multiple Transactional Time-Stamped Databases	157
9.4	Problem Statement	159
9.5	Clustering Items	161
9.5.1	Finding the Best Non-trivial Partition	163
9.5.2	Finding a Best Class	166
9.6	Experiments	168
9.7	Conclusions	171
	References.	172
10	Synthesizing Some Extreme Association Rules from Multiple Databases.	173
10.1	Introduction	173
10.2	Some Extreme Types of Association Rule in Multiple Databases	175
10.3	Problem Statement	177
10.4	An Extended Model of Local Pattern Analysis for Synthesizing Global Patterns	178
10.5	Related Work	180
10.6	Synthesizing an Association Rule	181
10.6.1	Design of Algorithm	183
10.6.2	Error Calculation	187
10.7	Experiments	188
10.7.1	Results of Experimental Studies	190
10.7.2	Comparison with Existing Algorithm.	191
10.8	Conclusions	194
	References.	194
11	Clustering Local Frequency Items in Multiple Data Sources	197
11.1	Introduction	197
11.2	Related Work	199
11.2.1	Measures of Association	199
11.2.2	Multi-database Mining Techniques	200
11.2.3	Clustering Techniques	203
11.3	Problem Statement	204
11.4	Synthesizing Support of an Itemset	205

11.5	Clustering Local Frequency Items	208
11.5.1	Finding the Best Non-trivial Partition	211
11.5.2	Error Analysis	215
11.6	Experimental Results.	216
11.6.1	Overall Output	218
11.6.2	Synthesis of High Frequency Itemsets	219
11.6.3	Error Quantification	221
11.6.4	Average Error Versus γ	222
11.6.5	Average Error Versus α	224
11.6.6	Clustering Time Versus Number of Databases	225
11.6.7	Comparison with Existing Technique	227
11.7	Conclusions	229
	References.	229
12	Mining Patterns of Select Items in Different Data Sources	233
12.1	Introduction	233
12.2	Mining Global Patterns of Select Items	235
12.3	Overall Association Between Two Items in a Database	237
12.4	An Application: Study of Select Items in Multiple Databases Through Grouping	240
12.4.1	Properties of Different Measures.	242
12.4.2	Grouping Frequent Items	244
12.4.3	Experiments	248
12.5	Related Work	251
12.6	Conclusions	253
	References.	253
13	Mining Calendar-Based Periodic Patterns in Time-Stamped Data	255
13.1	Introduction	255
13.2	Related Work	258
13.3	Calendar-Based Periodic Patterns	259
13.3.1	Extending Certainty Factor.	260
13.3.2	Extending Certainty Factor with Respect to Other Intervals	263
13.4	Mining Calendar-Based Periodic Patterns.	265
13.4.1	Improving Mining Calendar-Based Periodic Patterns.	265
13.4.2	Data Structure	266
13.4.3	A Modified Algorithm.	267

13.5	Experimental Studies	270
13.5.1	Selection of <i>Mininterval</i> and <i>Maxgap</i>	273
13.5.2	Selection of <i>Minsupp</i>	276
13.5.3	Performance Analysis	277
13.6	Conclusions	279
	References	280
14	Measuring Influence of an Item in Time-Stamped Databases	283
14.1	Introduction	283
14.2	Association Between Two Itemsets	285
14.3	Concept of Influence	286
14.3.1	Influence of an Itemset on Another Itemset	287
14.3.2	Properties of Influence Measures	288
14.3.3	Influence of an Item on a Set of Specific Items	289
14.3.4	Motivation	290
14.4	Problem Statement	291
14.5	Related Work	294
14.6	Design of Algorithms	294
14.6.1	Designing Algorithm for Measuring Overall Influence of an Item on Another Item	295
14.6.2	Designing Algorithm for Measuring Overall Influence of an Item on Each of the Specific Items	296
14.6.3	Designing Algorithm for Identifying Top Influential Items on a Set of Specific Items	296
14.7	Experiments	297
14.8	Conclusions	303
	References	303
15	Clustering Multiple Databases Induced by Local Patterns	305
15.1	Introduction	305
15.2	Problem Statement	308
15.3	Related Work	308
15.4	Clustering Databases	309
15.4.1	Finding the Best Non-trivial Partition	321
15.4.2	Efficiency of Clustering Technique	324
15.5	Experiments	328
15.6	Conclusions	331
	References	331
16	Enhancing Quality of Patterns in Multiple Related Databases	333
16.1	Introduction	333
16.2	Related Work	336
16.3	Simple Bit Vector (SBV) Coding	338

- 16.3.1 Dealing with Databases Containing Large
Number of Items 340
 - 16.4 Antecedent-Consequent Pair (ACP) Coding 341
 - 16.4.1 Indexing Rule Codes 345
 - 16.4.2 Storing Rulebases in Secondary Memory 349
 - 16.4.3 Space Efficiency of Our Approach 350
 - 16.5 Experiments 352
 - 16.6 Conclusions 355
 - References. 355
- 17 Concluding Remarks 357**
 - 17.1 Mass Storage Systems 357
 - 17.2 Some Emerging Domains in KDD 358
 - 17.3 New Patterns and Their Recognitions 360
 - 17.4 New Measures 361
 - 17.5 Algorithms and Techniques 362
 - 17.6 Challenges Ahead 364
 - References. 366
- Index 369**

Advances in Knowledge Discovery in Databases

Adhikari, A.; Adhikari, J.

2015, XV, 370 p. 136 illus., Hardcover

ISBN: 978-3-319-13211-2