

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation	1
1.2	Distributed Data Mining	3
1.3	Multi-Database Mining Approaches	4
1.3.1	Local Pattern Analysis	5
1.3.2	Sampling	6
1.3.3	Re-Mining	6
1.4	Pre-Processing of Databases	6
1.4.1	Preparation of Data Warehouses	7
1.4.2	Temporal Aggregation	8
1.4.3	Partitioning Database	8
1.4.4	Database Thinning	9
1.4.5	Ordering of Databases	9
1.4.6	Selection of Databases	9
1.5	Patterns and Associations	10
1.6	Related Studies	13
1.7	Experimental Settings	14
1.8	Conclusions	14
	References	16
<b>2</b>	<b>Synthesizing Different Extreme Association Rules from Multiple Databases</b>	<b>21</b>
2.1	Introduction	21
2.2	Some Extreme Types of Association Rule in Multiple Databases	23
2.3	Problem Statement	25
2.4	An Extended Model of Local Pattern Analysis for Synthesizing Global Patterns	25
2.5	Related Work	27
2.6	Synthesizing an Association Rule	29
2.6.1	Design of Algorithm	30
2.6.2	Error Calculation	34

2.7	Experiments . . . . .	35
2.7.1	Results of Experimental Studies . . . . .	37
2.7.2	Comparison with Existing Algorithm . . . . .	38
2.8	Conclusions . . . . .	40
	References . . . . .	41
<b>3</b>	<b>Clustering Items in Time-Stamped Databases</b>	
	<b>Induced by Stability . . . . .</b>	<b>43</b>
3.1	Introduction . . . . .	43
3.2	Related Work. . . . .	44
3.3	A Model of Mining Multiple Transactional Time-Stamped Databases. . . . .	45
3.4	Problem Statement . . . . .	47
3.5	Clustering Items . . . . .	49
3.5.1	Finding the Best Non-Trivial Partition . . . . .	50
3.5.2	Finding a Best Class . . . . .	54
3.6	Experiments. . . . .	55
3.7	Conclusions . . . . .	58
	References . . . . .	59
<b>4</b>	<b>Synthesizing Global Patterns in Multiple Large Data Sources . . .</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	Multi-Database Mining Using Local Pattern Analysis. . . . .	62
4.3	Generalized Multi-Database Mining Techniques. . . . .	63
4.3.1	Local Pattern Analysis . . . . .	63
4.3.2	Partition Algorithm. . . . .	64
4.3.3	IdentifyExPattern Algorithm . . . . .	64
4.3.4	RuleSynthesizing Algorithm. . . . .	64
4.4	Specialized Multi-Database Mining Techniques . . . . .	65
4.4.1	Mining Multiple Real Databases. . . . .	65
4.4.2	Mining Multiple Databases for the Purpose of Studying a Set of Items. . . . .	66
4.4.3	Study of Temporal Patterns in Multiple Databases . . .	66
4.5	Mining Multiple Databases Using Pipelined Feedback Technique . . . . .	66
4.5.1	Algorithm Design. . . . .	68
4.6	Error Evaluation. . . . .	68
4.7	Experiments. . . . .	69
4.8	Conclusions . . . . .	73
	References . . . . .	73

<b>5</b>	<b>Clustering Local Frequency Items in Multiple Data Sources . . . .</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Related Work. . . . .	77
5.2.1	Measures of Association . . . . .	77
5.2.2	Multi-Database Mining Techniques. . . . .	78
5.2.3	Clustering Techniques. . . . .	81
5.3	Problem Statement . . . . .	82
5.4	Synthesizing Support of an Itemset. . . . .	83
5.5	Clustering Local Frequency Items . . . . .	86
5.5.1	Finding the Best Non-Trivial Partition . . . . .	89
5.5.2	Error Analysis . . . . .	93
5.6	Experimental Results . . . . .	94
5.6.1	Overall Output . . . . .	96
5.6.2	Synthesis of High Frequency Itemsets. . . . .	97
5.6.3	Error Quantification . . . . .	99
5.6.4	Average Error Versus $\gamma$ . . . . .	100
5.6.5	Average Error Versus $\alpha$ . . . . .	101
5.6.6	Clustering Time Versus Number of Databases . . . . .	103
5.6.7	Comparison with Existing Technique . . . . .	104
5.7	Conclusions . . . . .	106
	References . . . . .	107
<b>6</b>	<b>Mining Patterns of Select Items in Different Data Sources . . . . .</b>	<b>109</b>
6.1	Introduction . . . . .	109
6.2	Mining Global Patterns of Select Items. . . . .	111
6.3	Overall Association Between Two Items in a Database. . . . .	113
6.4	An Application: Study of Select Items in Multiple Databases Through Grouping. . . . .	116
6.4.1	Properties of Different Measures . . . . .	118
6.4.2	Grouping Frequent Items. . . . .	120
6.4.3	Experiments. . . . .	123
6.5	Related Work. . . . .	127
6.6	Conclusions . . . . .	128
	References . . . . .	128
<b>7</b>	<b>Synthesizing Global Exceptional Patterns in Different Data Sources . . . . .</b>	<b>131</b>
7.1	Introduction . . . . .	131
7.2	Exceptional Patterns in Multiple Data Sources. . . . .	133
7.3	Problem Statement . . . . .	139
7.4	Related Work. . . . .	139
7.5	Synthesizing Support of an Itemset. . . . .	140
7.6	Synthesizing Type II Global Exceptional Itemsets . . . . .	141
7.7	Error Calculation . . . . .	145

7.8	Experiments . . . . .	146
7.8.1	Comparison with the Existing Algorithm. . . . .	151
7.9	Conclusions . . . . .	153
	References . . . . .	153
<b>8</b>	<b>Mining Icebergs in Different Time-Stamped Data Sources . . . . .</b>	<b>157</b>
8.1	Introduction . . . . .	157
8.2	Related Work. . . . .	159
8.3	Notches in Sales Series . . . . .	160
8.3.1	Identifying Notches. . . . .	162
8.4	Generalized Notch . . . . .	162
8.5	Iceberg Notch . . . . .	163
8.6	Sales Series . . . . .	164
8.6.1	Another View at Sales Series. . . . .	164
8.6.2	Other Applications of Icebergs. . . . .	165
8.7	Mining Icebergs in Time-Stamped Databases. . . . .	166
8.7.1	Non-Incremental Approach . . . . .	166
8.7.2	Incremental Approach . . . . .	169
8.8	Significant Year . . . . .	171
8.9	Experimental Studies . . . . .	172
8.10	Conclusions . . . . .	180
	References . . . . .	180
<b>9</b>	<b>Mining Calendar-Based Periodic Patterns in Time-Stamped Data . . . . .</b>	<b>183</b>
9.1	Introduction . . . . .	183
9.2	Related Work. . . . .	186
9.3	Calendar-Based Periodic Patterns . . . . .	187
9.3.1	Extending Certainty Factor . . . . .	188
9.3.2	Extending Certainty Factor with Respect to Other Intervals . . . . .	191
9.4	Mining Calendar-Based Periodic Patterns . . . . .	192
9.4.1	Improving Mining Calendar-Based Periodic Patterns. . . . .	193
9.4.2	Data Structure . . . . .	193
9.4.3	A Modified Algorithm . . . . .	195
9.5	Experimental Studies . . . . .	198
9.5.1	Selection of Mininterval and Maxgap . . . . .	202
9.5.2	Selection of Minsupp . . . . .	204
9.5.3	Performance Analysis . . . . .	205
9.6	Conclusions . . . . .	207
	References . . . . .	207

<b>10</b>	<b>Measuring Influence of an Item in Time-Stamped Databases . . . .</b>	<b>209</b>
10.1	Introduction . . . . .	209
10.2	Association Between Two Itemsets. . . . .	211
10.3	Concept of Influence. . . . .	212
10.3.1	Influence of an Itemset on Another Itemset . . . . .	213
10.3.2	Properties of Influence Measures . . . . .	214
10.3.3	Influence of an Item on a Set of Specific Items . . . . .	215
10.3.4	Motivation . . . . .	216
10.4	Problem Statement . . . . .	218
10.5	Related Work. . . . .	219
10.6	Design of Algorithms . . . . .	219
10.6.1	Designing Algorithm for Measuring Overall Influence of an Item on Another Item. . . . .	220
10.6.2	Designing Algorithm for Measuring Overall Influence of an Item on Each of the Specific Items. . . . .	221
10.6.3	Designing Algorithm for Identifying Top Influential Items on a Set of Specific Items. . . . .	221
10.7	Experiments. . . . .	222
10.8	Conclusions . . . . .	228
	References . . . . .	228
<b>11</b>	<b>Summary and Conclusions . . . . .</b>	<b>231</b>
11.1	Changing Scenarios . . . . .	231
11.2	Summary of Chapters . . . . .	232
11.3	Selected Open Problems and Challenges . . . . .	234
11.4	Conclusions . . . . .	234
	References . . . . .	235
	<b>Index . . . . .</b>	<b>237</b>

Data Analysis and Pattern Recognition in Multiple  
Databases

Adhikari, A.; Adhikari, J.; Pedrycz, W.

2014, XV, 238 p. 97 illus., Hardcover

ISBN: 978-3-319-03409-6