

Knowledge Discovery for Business Information Systems

Edited by

Witold Abramowicz

The Poznan University of Economics, Poland

Jozef Zurada

University of Louisville, U.S.A.



KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Contents

PREFACE	xi
FOREWORD	xiii
LIST OF CONTRIBUTORS	xv
Chapter 1 INFORMATION FILTERS SUPPLYING DATA WAREHOUSES WITH BENCHMARKING INFORMATION	1
<i>Witold Abramowicz, PawelJan Kalczyhski, Krzysztof Wecel</i>	
1. Introduction	1
2. Data Warehouses	2
3. The HyperSDI System	4
4. User Profiles in the HyperSDI System	11
5. Building Data Warehouse Profiles	11
6. Techniques for Improving Profiles	18
7. Implementation Notes	22
8. Conclusions	25
References	
Chapter 2 PARALLEL MINING OF ASSOCIATION RULES	29
<i>David Cheung, Sau Dan Lee</i>	
1. Introduction	29
2. Parallel Mining of Association Rules	32
3. Pruning Techniques and The FPM Algorithm	33
4. Metrics for Data Skewness and Workload Balance	39
5. Partitioning of the Database	48
6. Experimental Evaluation of the Partitioning Algorithms	56
7. Discussions	62
8. Conclusions	64
References	65

Chapter 3	UNSUPERVISED FEATURE RANKING AND SELECTION	67
	<i>Manoranjan Dash, Huan Liu, Jun Yao</i>	
	1. Introduction	67
	2. Basic Concepts and Possible Approaches	69
	3. An Entropy Measure for Continuous and Nominal Data Types	72
	4. Algorithm to Find Important Variables	75
	5. Experimental Studies	76
	6. Clustering Using SUD	80
	7. Discussion and Conclusion	82
	References	84
 Chapter 4	 APPROACHES TO CONCEPT BASED EXPLORATION OF INFORMATION RESOURCES	 89
	<i>Hele-Mai Haav, Jergen Fischer Nilsson</i>	
	1. Introduction	89
	2. Conceptual Taxonomies	91
	3. Ontology Driven Concept Retrieval	99
	4. Search based on formal concept analysis	104
	5. Conclusion	109
	Acknowledgements	109
	References	109
 Chapter 5	 HYBRID METHODOLOGY OF KNOWLEDGE DISCOVERY FOR BUSINESS INFORMATION	 111
	<i>Zdzislaw S. Hippe</i>	
	1. Introduction	111
	2. Present Status of Data Mining	113
	3. Experiments with Mining Regularities from Data	118
	4. Discussion	125
	Acknowledgements	126
	References	126
 Chapter 6	 FUZZY LINGUISTIC SUMMARIES OF DATABASES FOR AN EFFICIENT BUSINESS DATA ANALYSIS AND DECISION SUPPORT	 129
	<i>Janusz Kacprzyk, Ronald R. Yager and Siawomir Zadrozny</i>	
	1. Introduction	129
	2. Idea of Linguistic Summaries Using Fuzzy Logic with Linguistic Quantifiers	131
	3. On Other Validity Criteria	134
	4. Derivation of Linguistic Summaries via a Fuzzy Logic Based Database Querying Interface	140
	5. Implementation for a Sales Database at a Computer Retailer	147

6. Concluding Remarks	150
References	150
Chapter 7	INTEGRATING DATA SOURCES USING A STANDARDIZED GLOBAL DICTIONARY
	153
<i>Ramon Lawrence and Ken Barker</i>	
1. Introduction	154
2. Data Semantics and the Integration Problem	154
3. Previous work	156
4. The Integration Architecture	157
5. The Global Dictionary	160
6. The Relational Integration Model	164
7. Special Cases of Integration	169
8. Applications to the WWW	171
9. Future Work and Conclusions	171
References	172
Chapter 8	MAINTENANCE OF DISCOVERED ASSOCIATION RULES
	173
<i>Sau Dan Lee, David Cheung</i>	
1. Introduction	173
2. Problem Description	176
3. The FUP Algorithm for the Insertion Only Case	179
4. The FUP Algorithm for the Deletions Only Case	183
5. The FUP ₂ Algorithm for the General Case	189
6. Performance Studies	194
7. Discussions	204
8. Conclusions	208
Notes	209
References	209
Chapter 9	MULTIDIMENSIONAL BUSINESS PROCESS ANALYSIS WITH THE PROCESS WAREHOUSE
	211
<i>Beate List, Josef Schiefer, A Min Tjoa, Gerald Quirchmayr</i>	
1. Introduction	211
2. Related Work	213
3. Goals of the Data Warehouse Approach	215
4. Data Source	216
5. Basic Process Warehouse Components Representing Business Process Analysis Requirements	216
6. Data Model and Analysis Capabilities	219
7. Conclusion and Further Research	225
References	225

Chapter 10	AMALGAMATION OF STATISTICS AND DATA MINING TECHNIQUES: EXPLORATIONS IN CUSTOMER LIFETIME VALUE MODELING	229
	<i>D. R. Mani, James Drew, Andrew Betz and Piew Datta</i>	
	1. Introduction	229
	2. Statistics and Data Mining Techniques: A Characterization	231
	3. Lifetime Value (LTV) Modeling	232
	4. Customer Data for LTV Tenure Prediction	234
	5. Classical Statistical Approaches to Survival Analysis	235
	6. Neural Networks for Survival Analysis	239
	7. From Data Models to Business Insight	244
	8. Conclusion: The Amalgamation of Statistical and Data Mining Techniques	247
	References	249
 Chapter 11	 ROBUST BUSINESS INTELLIGENCE SOLUTIONS	 251
	<i>Jan Mrazek</i>	
	1. Introduction	251
	2. Business Intelligence Architecture	252
	3. Data Transformation	258
	4. Data Modelling	260
	5. Integration Of Data Mining	268
	6. Conclusion	272
	References	273
 Chapter 12	 THE ROLE OF GRANULAR INFORMATION IN KNOWLEDGE DISCOVERY IN DATABASES	 275
	<i>Witold Pedrycz</i>	
	1. Introduction	276
	2. Granulation of information	277
	3. The development of data-justifiable information granules	284
	4. Building associations in databases	287
	5. From associations to rules in databases	292
	6. The construction of rules in data mining	293
	7. Properties of rules induced by associations	298
	8. Detailed computations of the consistency of rules and its analysis	300
	9. Conclusions	303
	Acknowledgment	304
	References	304

Chapter 13	DEALING WITH DIMENSIONS IN DATA WAREHOUSING	307
	<i>Jaroslav Pokorny</i>	
	1. Introduction	308
	2. DW Modelling with Tables	310
	3. Dimensions	311
	4. Constellations	314
	5. Dimension Hierarchies with ISA-hierarchies	315
	6. Conclusions	323
	References	324
Chapter 14	ENHANCING THE KDD PROCESS IN THE RELATIONAL DATABASE MINING FRAMEWORK BY QUANTITATIVE EVALUATION OF ASSOCIATION RULES	325
	<i>Giuseppe Psaila</i>	
	1. Introduction	325
	2. The Relational Database Mining Framework	327
	3. The Evaluate Rule Operator	331
	4. Enhancing the Knowledge Discovery Process	344
	5. Conclusions and Future Work	348
	Notes	349
	References	349
Chapter 15	SPEEDING UP HYPOTHESIS DEVELOPMENT	351
	<i>Jorg A. Schlosser, Peter C. Lockemann, Matthias Gimbel</i>	
	1. Introduction	352
	2. Information Model	354
	3. The Execution Architecture of CITRUS	356
	4. Searching the Information Directory	359
	5. Documentation of the Process History	361
	6. Linking the Information Model with the Relational Model	362
	7. Generation of SQL Queries	364
	8. Automatic Materialization of Intermediate Results	368
	9. Experimental Results	369
	10. Utilizing Past Experience	371
	11. Related Work	372
	12. Concluding Remarks	374
	References	374

Chapter 16	SEQUENCE MINING IN DYNAMIC AND INTERACTIVE ENVIRONMENTS	377
	<i>Srinivasan Parthasarathy, Mohammed J. Zaki, Mitsunori Ogihara, Sandhya Dwarkadas</i>	
	1. Introduction	378
	2. Problem Formulation	379
	3. The SPADE Algorithm	382
	4. Incremental Mining Algorithm	384
	5. Interactive Sequence Mining	388
	6. Experimental Evaluation	390
	7. Related Work	394
	8. Conclusions	395
	Acknowledgements	395
	References	395
 Chapter 17	 INVESTIGATION OF ARTIFICIAL NEURAL NETWORKS FOR CLASSIFYING LEVELS OF FINANCIAL DISTRESS OF FIRMS: THE CASE OF AN UNBALANCED TRAINING SAMPLE	 397
	<i>Jozef Zurada, Benjamin P. Foster, Terry J. Ward</i>	
	1. Introduction	398
	2. Motivation and Literature Review	399
	3. Logit Regression, Neural Network, and Principal Component Analysis Fundamentals	402
	4. Research Methodology	409
	5. Discussion of the Results	415
	6. Conclusions and Future Research Directions	420
	Appendix - Neural Network Toolbox	421
	References	423
 INDEX		 425