

Contents

1. Importance of Multi-database Mining	1
1.1 Introduction	1
1.2 Role of Multi-database Mining in Real-world Applications	2
1.3 Multi-database Mining Problems	4
1.4 Differences Between Mono- and Multi-database Mining	6
1.4.1 Features of Data in Multi-databases	6
1.4.2 Features of Patterns in Multi-databases	8
1.5 Evolution of Multi-database Mining	9
1.6 Limitations of Previous Techniques	12
1.7 Process of Multi-database Mining	14
1.7.1 Description of Multi-database Mining	14
1.7.2 Practical Issues in the Process	16
1.8 Features of the Defined Process	20
1.9 Major Contributions of This Book	23
1.10 Organization of the Book	24
2. Data Mining and Multi-database Mining	27
2.1 Introduction	27
2.2 Knowledge Discovery in Databases	28
2.2.1 Processing Steps of KDD	28
2.2.2 Data Pre-processing	30
2.2.3 Data Mining	31
2.2.4 Post Data Mining	33
2.2.5 Applications of KDD	34
2.3 Association Rule Mining	36
2.4 Research into Mining Mono-databases	41
2.5 Research into Mining Multi-databases	51
2.5.1 Parallel Data Mining	51
2.5.2 Distributed Data Mining	52
2.5.3 Application-dependent Database Selection	58
2.5.4 Peculiarity-oriented Multi-database Mining	59
2.6 Summary	61

3. Local Pattern Analysis	63
3.1 Introduction	63
3.2 Previous Multi-database Mining Techniques	64
3.3 Local Patterns	65
3.4 Local Instance Analysis Inspired by Competition in Sports...	67
3.5 The Structure of Patterns in Multi-database Environments	70
3.6 Effectiveness of Local Pattern Analysis	73
3.7 Summary	74
4. Identifying Quality Knowledge	75
4.1 Introduction	75
4.2 Problem Statement	76
4.2.1 Problems Faced by Traditional Multi-database Mining	76
4.2.2 Effectiveness of Identifying Quality Data	78
4.2.3 Needed Concepts	80
4.3 Nonstandard Interpretation	82
4.4 Proof Theory	88
4.5 Adding External Knowledge	91
4.6 The Use of the Framework	95
4.6.1 Applying to Real-world Applications	95
4.6.2 Evaluating Veridicality	96
4.7 Summary	100
5. Database Clustering	103
5.1 Introduction	103
5.2 Effectiveness of Classifying	104
5.3 Classifying Databases	107
5.3.1 Features in Databases	107
5.3.2 Similarity Measurement	108
5.3.3 Relevance of Databases and Classification	113
5.3.4 Ideal Classification and Goodness Measurement	115
5.4 Searching for a Good Classification	120
5.4.1 The First Step: Generating a Classification	121
5.4.2 The Second Step: Searching for a Good Classification	123
5.5 Algorithm Analysis	127
5.5.1 Procedure <i>GreedyClass</i>	127
5.5.2 Algorithm <i>GoodClass</i>	129
5.6 Evaluation of Application-independent Database Classification	130
5.6.1 Dataset Selection	130
5.6.2 Experimental Results	131
5.6.3 Analysis	134
5.7 Summary	135

6. Dealing with Inconsistency	137
6.1 Introduction	137
6.2 Problem Statement	138
6.3 Definitions of Formal Semantics	139
6.4 Weighted Majority	143
6.5 Mastering Local Pattern Sets	146
6.6 Examples of Synthesizing Local Pattern Sets	148
6.7 A Syntactic Characterization	150
6.8 Summary	155
7. Identifying High-vote Patterns	157
7.1 Introduction	157
7.2 Illustration of High-vote Patterns	158
7.3 Identifying High-vote Patterns	161
7.4 Algorithm Design	163
7.4.1 Searching for High-vote Patterns	164
7.4.2 Identifying High-vote Patterns: An Example	165
7.4.3 Algorithm Analysis	167
7.5 Identifying High-vote Patterns Using a Fuzzy Logic Controller	168
7.5.1 Needed Concepts in Fuzzy Logic	168
7.5.2 System Analysis	170
7.5.3 Setting Membership Functions for Input and Output Variables	171
7.5.4 Setting Fuzzy Rules	172
7.5.5 Fuzzification	174
7.5.6 Inference and Rule Composition	174
7.5.7 Defuzzification	176
7.5.8 Algorithm Design	177
7.6 High-vote Pattern Analysis	178
7.6.1 Normal Distribution	178
7.6.2 The Procedure of Clustering	179
7.7 Suggested Patterns	183
7.8 Summary	183
8. Identifying Exceptional Patterns	185
8.1 Introduction	185
8.2 Interesting Exceptional Patterns	186
8.2.1 Measuring the Interestingness	186
8.2.2 Behavior of Interest Measurements	189
8.3 Algorithm Design	189
8.3.1 Algorithm Design	189
8.3.2 Identifying Exceptions: An Example	192
8.3.3 Algorithm Analysis	193
8.4 Identifying Exceptions with a Fuzzy Logic Controller	195
8.5 Summary	195

9. Synthesizing Local Patterns by Weighting	197
9.1 Introduction	197
9.2 Problem Statement	198
9.3 Synthesizing Rules by Weighting	200
9.3.1 Weight of Evidence	200
9.3.2 Solving Weights of Databases	201
9.3.3 Algorithm Design	205
9.4 Improvement of Synthesizing Model	206
9.4.1 Effectiveness of Rule Selection	206
9.4.2 Process of Rule Selection	208
9.4.3 Optimized Algorithm	210
9.5 Algorithm Analysis	211
9.5.1 Procedure <i>RuleSelection</i>	211
9.5.2 Algorithm <i>RuleSynthesizing</i>	212
9.6 Summary	213
10. Conclusions and Future Work	215
10.1 Conclusions	215
10.2 Future Work	218
References	221
Subject Index	231



<http://www.springer.com/978-1-85233-703-2>

Knowledge Discovery in Multiple Databases

Zhang, S.; Zhang, C.; Wu, X.

2004, XII, 233 p., Hardcover

ISBN: 978-1-85233-703-2