

Contents

Part I Fundamental Concepts

1	Introduction	3
1.1	Privacy Preserving Data Mining	5
1.2	Association Rule Hiding	5
1.3	Research Challenges	7
1.4	Summary	8
2	Background	9
2.1	Terminology and Preliminaries	9
2.2	Problem Formulation and Statement	12
2.2.1	Goals of Association Rule Hiding Methodologies	13
2.2.2	Problem Statement	14
3	Classes of Association Rule Hiding Methodologies	17
3.1	Classification Dimensions	17
3.2	Classes of Association Rule Hiding Algorithms	19
4	Other Knowledge Hiding Methodologies	21
4.1	Classification Rule Hiding	21
4.2	Privacy Preserving Clustering	22
4.3	Sequence Hiding	23
5	Summary	25

Part II Heuristic Approaches

6	Distortion Schemes	29
7	Blocking Schemes	35
8	Summary	37

Part III Border Based Approaches

- 9 Border Revision for Knowledge Hiding** 41
- 10 BBA Algorithm** 47
 - 10.1 Hiding Goals 48
 - 10.2 Hiding Process 49
 - 10.2.1 Weighing Border Itemsets 50
 - 10.2.2 Hiding a Sensitive Itemset 51
 - 10.2.3 Order of Hiding Itemsets 51
 - 10.3 Algorithm 52
- 11 Max–Min Algorithms** 53
 - 11.1 Hiding a Sensitive Itemset 54
 - 11.2 Order of Hiding Itemsets 55
 - 11.3 Max–Min 1 Algorithm 55
 - 11.4 Max–Min 2 Algorithm 56
- 12 Summary** 59

Part IV Exact Hiding Approaches

- 13 Menon’s Algorithm** 63
 - 13.1 Exact Part 64
 - 13.1.1 CSP Formulation 64
 - 13.1.2 CSP Decomposition 65
 - 13.2 Heuristic Part 69
- 14 Inline Algorithm** 71
 - 14.1 Privacy Model 72
 - 14.2 Solution Methodology 74
 - 14.2.1 Problem Size Minimization 74
 - 14.2.2 Reduction to CSP and the BIP solution 75
 - 14.2.3 An Example 77
 - 14.3 Experiments and Results 79
 - 14.3.1 Datasets 80
 - 14.3.2 Evaluation Methodology 80
 - 14.3.3 Experimental Results 81
 - 14.3.4 Discussion on the Efficiency of the Inline Algorithm 81
- 15 Two–Phase Iterative Algorithm** 83
 - 15.1 Background 83
 - 15.2 Removal of Constraints from Infeasible CSPs 85
 - 15.3 An Example 86
 - 15.4 Experimental Evaluation 89

- 16 Hybrid Algorithm** 93
 - 16.1 Knowledge Hiding Formulation 94
 - 16.1.1 Hybrid Hiding Methodology 94
 - 16.1.2 A Running Example 96
 - 16.2 Main Issues Pertaining to the Hiding Methodology 96
 - 16.2.1 Size of Database Extension 96
 - 16.2.2 Optimal Solutions in the Hybrid Methodology 98
 - 16.2.3 Revision of the Borders 99
 - 16.2.4 Problem Size Reduction 100
 - 16.2.5 Handling Suboptimality 101
 - 16.3 Hybrid Solution Methodology 102
 - 16.3.1 Problem Size Minimization 102
 - 16.3.2 Adjusting the Size of the Extension 104
 - 16.3.3 Formulation and Solution of the CSP 105
 - 16.3.4 Minimum Extension and Validity of Transactions 105
 - 16.3.5 Treatment of Suboptimality in Hiding Solutions 107
 - 16.3.6 Continuing the Running Example 108
 - 16.4 A Partitioning Approach to Improve Scalability 110
 - 16.4.1 Partitioning Methodology 110
 - 16.4.2 An Example 111
 - 16.5 Experimental Evaluation 112

- 17 Parallelization Framework for Exact Hiding** 119
 - 17.1 The Parallelization Framework 119
 - 17.1.1 Structural Decomposition of the CSP 120
 - 17.1.2 Decomposition of Large Independent Components 123
 - 17.1.3 Parallel Solving of the Produced CSPs 126
 - 17.2 Computational Experiments and Results 128

- 18 Quantifying the Privacy of Exact Hiding Algorithms** 131

- 19 Summary** 135

- Part V Epilogue**

- 20 Conclusions** 139

- 21 Roadmap to Future Work** 143

- References** 145

- Index** 149



<http://www.springer.com/978-1-4419-6568-4>

Association Rule Hiding for Data Mining

Gkoulalas-Divanis, A.; Verykios, V.S.

2010, XX, 138 p. 60 illus., Hardcover

ISBN: 978-1-4419-6568-4