

# Contents

<b>Certificate</b>	<b>iii</b>
<b>Declaration by the Candidate</b>	<b>v</b>
<b>Copyright Transfer Certificate</b>	<b>vii</b>
<b>Preface</b>	<b>xi</b>
<b>Acknowledgements</b>	<b>xv</b>
<b>Contents</b>	<b>xvi</b>
<b>List of Figures</b>	<b>xxiii</b>
<b>List of Tables</b>	<b>xxv</b>
<b>Symbols</b>	<b>xxxix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Issues and Challenges . . . . .	2
1.2.1 Constraint-based High Utility Itemsets Mining . . . . .	3
1.2.2 Top-k High Utility Itemsets Mining . . . . .	4
1.2.3 High Utility Itemsets Mining with Negative Utility Value . . . . .	5
1.2.4 Constraint-based High Utility Itemsets Mining with Negative Utility Value . . . . .	5
1.2.5 Closed High Utility Itemsets Mining with Negative Utility Value .	6
1.3 Contributions . . . . .	7

1.3.1	Constraint-based High Utility Itemsets Mining . . . . .	7
1.3.2	Top-k High Utility Itemsets Mining . . . . .	8
1.3.3	High Utility Itemsets Mining with Negative Utility Value . . . . .	8
1.3.4	Constraint-based High Utility Itemsets Mining with Negative Utility Value . . . . .	9
1.3.5	Closed High Utility Itemsets Mining with Negative Utility Value .	10
1.4	Organization . . . . .	10
<b>2</b>	<b>Literature Review</b>	<b>15</b>
2.1	High Utility Itemset Mining . . . . .	15
2.2	Constraint-based High Utility Itemset Mining . . . . .	22
2.3	Top-k High Utility Itemset Mining . . . . .	23
2.3.1	Two-Phase Algorithms . . . . .	23
2.3.2	One-Phase Algorithms . . . . .	25
2.4	Closed High Utility Itemset Mining . . . . .	26
2.4.1	Level-wise Mining Algorithms . . . . .	27
2.4.2	Tree-based Mining Algorithms . . . . .	28
2.4.3	Utility-list based Mining Algorithms . . . . .	29
2.4.4	Discussion . . . . .	31
2.5	High Utility Itemset Mining with Negative Utility Value . . . . .	31
2.5.1	Properties to Handle Negative Utility Items . . . . .	33
2.5.2	Level-wise Mining Algorithms . . . . .	37
2.5.3	Tree-based Mining Algorithms . . . . .	38
2.5.4	Utility-list based Mining Algorithms . . . . .	39
2.5.5	Discussion . . . . .	42
2.6	Summary . . . . .	43
<b>3</b>	<b>High Utility Itemsets Mining Considering Length Constraints</b>	<b>45</b>
3.1	EHIL Algorithm . . . . .	48
3.1.1	The Search Space . . . . .	49
3.1.2	Efficient Dataset Scanning Techniques . . . . .	50
3.1.2.1	Dataset Scanning using Projection . . . . .	50
3.1.2.2	Dataset Scanning using Merging . . . . .	50
3.1.3	Pruning Strategies . . . . .	52
3.1.3.1	Prune search space using Revised Local Utility . . . . .	52
3.1.3.2	Prune search space using Revised Sub-tree Utility . . . . .	53
3.1.3.3	Pruning using Length Constraints . . . . .	54
3.1.4	Calculate Upper Bounds using Utility Array . . . . .	55
3.1.5	The Proposed Algorithm . . . . .	57
3.1.6	An Illustrative Example . . . . .	59
3.2	Experimental Results . . . . .	60
3.2.1	Runtime Performance Comparison . . . . .	61

3.2.2	Memory Performance Comparison . . . . .	62
3.2.3	Relative Runtime and Memory Comparison Analysis . . . . .	65
3.2.3.1	Relative runtime comparison . . . . .	66
3.2.3.2	Relative memory comparison . . . . .	67
3.2.3.3	Relative best, average and minimum comparison . . . . .	68
3.2.4	Scalability . . . . .	69
3.3	Summary . . . . .	69
<b>4</b>	<b>Top-k High Utility Itemsets Mining</b>	<b>71</b>
4.1	Preliminaries and Problem Definition . . . . .	71
4.2	TKEH Algorithm . . . . .	74
4.2.1	The Search Space . . . . .	74
4.2.2	Concept of Co-occurrence Structure . . . . .	75
4.2.3	Dataset Scanning Techniques . . . . .	77
4.2.3.1	Dataset Scanning using Projection . . . . .	77
4.2.3.2	Dataset Scanning using Transaction Merging . . . . .	77
4.2.4	Threshold Raising Strategies . . . . .	78
4.2.4.1	RIU strategy . . . . .	79
4.2.4.2	CUD strategy . . . . .	79
4.2.4.3	COV strategy . . . . .	80
4.2.5	Pruning Strategies . . . . .	80
4.2.5.1	Prune search space using <i>EUCP</i> . . . . .	81
4.2.5.2	Prune search space using Sub-tree Utility . . . . .	81
4.2.6	Calculate Upper Bounds using Utility Array . . . . .	81
4.2.7	Main Procedure of TKEH . . . . .	83
4.3	Performance Evaluations . . . . .	85
4.3.1	Dense Datasets . . . . .	86
4.3.2	Sparse Datasets . . . . .	86
4.3.3	Memory Usage . . . . .	88
4.3.4	Scalability . . . . .	90
4.4	Discussion . . . . .	90
4.5	Summary . . . . .	91
<b>5</b>	<b>High Utility Itemsets Mining with Negative Utility Value</b>	<b>93</b>
5.1	EHIN Algorithm . . . . .	95
5.1.1	The Search Space . . . . .	95
5.1.2	Efficient Dataset Scanning Techniques . . . . .	96
5.1.2.1	Dataset Scanning using Projection . . . . .	96
5.1.2.2	Dataset Scanning using Transaction Merging . . . . .	97
5.1.3	Pruning Strategies . . . . .	99
5.1.3.1	Prune search space using redefined Local Utility . . . . .	100
5.1.3.2	Prune search space using redefined sub-tree Utility . . . . .	100

5.1.4	Calculate Upper Bounds using Utility Array . . . . .	102
5.1.5	An Illustrative Example . . . . .	107
5.2	Performance Evaluation . . . . .	108
5.2.1	Experiments on Dense Datasets . . . . .	111
5.2.2	Experiments on Sparse Datasets . . . . .	111
5.2.3	Memory Usage . . . . .	112
5.2.4	Relative Runtime and Memory Comparison Analysis . . . . .	114
5.2.4.1	Relative runtime analysis . . . . .	114
5.2.4.2	Relative memory usage analysis . . . . .	115
5.2.4.3	Relative runtime analysis on best, average and minimum case . . . . .	115
5.2.4.4	Relative memory usage analysis on best, average and minimum case . . . . .	117
5.2.5	Scalability . . . . .	117
5.3	Summary . . . . .	118
<b>6</b>	<b>High Utility Itemsets Mining with Negative Utility Value and Length Constraints</b>	<b>119</b>
6.1	The Proposed Algorithm . . . . .	122
6.1.1	Efficient Dataset Scanning Techniques . . . . .	123
6.1.2	Pruning Strategies . . . . .	124
6.1.2.1	Prune search space using Redefined Sub-tree Utility . . . . .	124
6.1.2.2	Pruning using Length Constraints . . . . .	126
6.1.3	Calculate Upper Bounds using Utility Array . . . . .	126
6.1.4	EHNL Algorithm . . . . .	127
6.1.5	An Illustrative Example . . . . .	131
6.2	Experimental Results . . . . .	134
6.2.1	Effect of Techniques . . . . .	140
6.2.2	Memory Usage . . . . .	143
6.2.3	Scalability Experiments . . . . .	145
6.2.4	Discussion . . . . .	147
6.3	Summary . . . . .	147
<b>7</b>	<b>Closed High Utility Itemsets Mining with Negative Utility Value</b>	<b>149</b>
7.1	The Proposed Algorithm . . . . .	150
7.1.1	Efficient Dataset Scanning Techniques . . . . .	150
7.1.2	Pruning Non-HUIs . . . . .	151
7.1.2.1	Prune search space using Redefined Sub-tree Utility . . . . .	152
7.1.3	Calculate Upper Bounds using Utility Array . . . . .	153
7.1.3.1	Calculating $RTWU$ of all items using $UA$ . . . . .	153
7.1.3.2	Calculating $RSU(\alpha)$ . . . . .	154
7.1.3.3	Calculating the $support(\alpha)$ : . . . . .	154

7.1.4	Closed HUIs Mining Strategies . . . . .	154
7.1.5	CHN Algorithm . . . . .	156
7.1.6	An Illustrative example . . . . .	159
7.2	Experimental Results . . . . .	161
7.2.1	Runtime Performance on Dense Datasets . . . . .	162
7.2.2	Runtime Performance on Sparse Datasets . . . . .	164
7.2.3	Memory Usages on Dense Datasets . . . . .	165
7.2.4	Memory Usages on Sparse Datasets . . . . .	166
7.2.5	Discussion . . . . .	167
7.2.6	Effect of Techniques . . . . .	168
7.2.7	Scalability . . . . .	169
7.3	Summary . . . . .	170
<b>8</b>	<b>Conclusion and Future Directions</b>	<b>171</b>
8.1	Conclusions . . . . .	171
8.2	Future Directions . . . . .	174
8.2.1	Mining High Utility Itemsets with Positive Utility only . . . . .	174
8.2.2	Mining High Utility Itemsets with Positive and Negative Utility . . . . .	175
8.2.3	Other High Utility Itemsets Problems . . . . .	175
<b>A</b>	<b>List of Publications</b>	<b>195</b>