

Contents

1	Introduction	1
1.1	Association Rule Mining (ARM)	2
1.2	Background Literature	2
1.3	With Candidate Generation	4
1.4	Without Candidate Generation Phase	5
1.5	Problem Statement	6
1.6	Motivation	6
1.7	Objective	7
1.8	Organization of the Thesis	8
2	Literature Survey	9
2.1	Different Set Representation Techniques available	9
2.2	Generation of Itemsets with Candidate Generation	10
2.2.1	Apriori	10
2.2.2	DIC	11
2.2.3	SETM	12
2.2.4	AIS	12
2.3	Generation of Itemsets without Candidate Generation	13

2.3.1	FP-growth	13
2.3.2	ECLAT	13
2.3.3	LCOFI	13
2.3.4	IR-Eclat	14
2.3.5	CAFP	14
2.3.6	Meta-PCP	14
2.4	Itemset Generation	15
2.4.1	Different Types of Set Representation used in ARM	15
2.5	Conclusion	16
3	Proposed Representation of Itemset	17
3.1	Motivation	17
3.2	Description of Proposed Representation	19
3.2.1	Dealing Numerical Attributes	21
3.2.2	Dealing Categorical Attributes	22
3.3	Theoretical Analysis	23
3.3.1	Effect of Attribute Values on the size of a single itemset	23
3.3.2	Existing Representations	23
3.3.3	Proposed Representation	25
3.4	Evaluation of the Performance of Proposed Representation for Apri- ori Algorithm	27
3.4.1	Apriori Algorithm	27
3.4.2	Incorporating the Itemset Representation for the Apriori and its variants	28
3.4.2.1	Itemset Generation	28

3.4.2.2	Proof of Correctness	29
3.4.2.3	Proof of Completeness	30
3.4.3	Incorporating the Searching Techniques for Itemsets in the Apriori Algorithm	31
3.4.3.1	Performance Analysis of Linear Search Technique used in Apriori Algorithm	32
3.4.3.2	Performance Analysis of Binary Search Techniques used in Apriori Algorithm	32
3.4.3.3	Conclusion	36
4	Graph Based Itemset Mining	40
4.1	Motivation for Graph Based Data Algorithm	40
4.2	Introduction to Graph Based Data Structure	41
4.3	Proposed Graph Based Algorithm	41
4.4	Experimental Results	42
4.5	Evaluation of the Proposed Algorithm	44
4.6	Proof of Correctness	50
4.7	Proof of Completeness	52
4.8	Discussion and Conclusion	52
5	Generate Rules from Mined Frequent Itemsets	53
5.1	Introduction	53
5.1.1	Agarwal's Algorithm	54
5.1.2	Srikant's 1 st Algorithm	54
5.1.3	Srikant's 2 nd Algorithm	55
5.1.4	NBG's Algorithm	55

Contents

5.2	Rule Generation with Existing Representation	55
5.3	Rule Generation with Proposed Representation	68
5.4	Discussion and Conclusion	69
6	Conclusion and Future Direction	72
6.1	Inferences	72
6.2	Future Work	72
A	Publication List	74