

Table of Contents

Acknowledgment

Abstract

1	Introduction	1
1.1	Data mining: an overview	1
1.2	Fundamentals of Association Rule Mining	3
1.3	Association Rule Mining : State-of-the-art	4
1.4	Motivation	4
1.5	Aim of the Work	4
1.6	Report Organization	5
2	Background of Association Rule Mining	6
2.1	Definition	6
2.2	Useful Concepts	6
2.3	Process	7
2.4	Example	8
2.5	Discussion	9
3.	Related Works	11
3.1	Frequent Itemset Mining	11
3.2	Association Rule Generation	17
3.3	Discussion	18
4.	Frequent Itemset Mining	19
4.1	Missing Value Prediction	19
4.2	Conversion of the given dataset into market basket dataset	20

4.3 Frequent Itemset Mining	21
4.3.1 Apriori Algorithm	22
4.4 Discussion	24
5. Association Rule Mining and Classification	25
5.1 Discovering association rules	25
5.2 Finding the set of Strong Rules	27
5.2.1 Extraction of small and non redundant set of rules.	27
5.2.2 Rule tuning	28
6. ARM in Cancer Data Classification	32
6.1 Dataset Description	33
6.2 Project Environment	35
6.3 Experimental Results	35
6.3.1 Summarized Experimental Result	36
6.3.2 Screen Shots	37
6.4 Discussion	40
7. ARM in Intrusion Data Classification	41
7.1 Dataset Description	42
7.2 Experimental Results	42
7.2.1 Summarized Experimental Results	43
7.2.2 Screen Shots	43
8. Conclusion and Future Work	
8.1 Conclusion	46
8.2 Future Work	46
References	48

List of Figures

Figure 1.1: Data mining as a part of knowledge discovery process	2
Figure 2.1: Generation of frequent itemsets from transactional data of <i>AllElectronics</i> branch	9
Figure 4.1: A portion of the breast cancer dataset	20
Figure 4.2: Converted market basket dataset	21
Figure 5.1: Flow diagram for extracting non- redundant set of asso- citation rules	30
Figure 5.2: Flow diagram for rule tuning	31
Figure 6.1: A portion of the frequent itemset generated from breast cancer dataset	37
Figure 6.2: A portion of association rules generated from breast cancer dataset(1)	37
Figure 6.3: A portion of association rules generated from breast cancer dataset(2)	38
Figure 6.4: A portion of strong rules after pruning (1)	38
Figure 6.5: A portion of strong rules after pruning (2)	39
Figure 6.6: Classification results	39
Figure 7.1: Description of 19 features of KDD 99 dataset	42
Figure 7.2: A portion of the generated frequent itemsets from the KDD 99 sample dataset	43
Figure 7.3: A portion of the generated rules	44
Figure 7.4: Association rules after pruning	44
Figure 7.5: Classification result	45

List of Tables

Table 2.1: Transactional data from <i>AllElectronics</i> branch	8
Table 3.1: Characteristics of different AM Techniques	18
Table 6.1: Experimental results of association mining in breast cancer dataset	36
Table 7.1: Experimental results of association mining in KDD 99 sample dataset	43