LIST OF FIGURES

Figure 1.1 An example for association rule mining through the discovery of frequent itemset. 2

Figure 1.2 An example database 2

Figure 2.1 The algorithms for frequent itemset mining can be categorized into four types. 6

Figure 2.2 Example. 6

Figure 2.3 Apriori algorithm. 8

Figure 2.4 Apriori-gen function in algorithm Apriori. 8

Figure 2.5 An example for Apriori. 9

Figure 2.6 The candidate 2-itemsets in Figure 2.5 is stored in this hash tree and the dark nodes or hash tables are traversed when scanning the transaction 200, {2, 3, 5}. 10

Figure 2.7 The traversal method for Apriori and Eclat. 11

Figure 2.8 Horizontal and vertical database layout. 12

Figure 2.9 An example for Eclat. 13

Figure 2.10 An example for Eclat that represent transacton-ID list as a bit matrix. 13

Figure 2.11 An example for FP-Tree construction. 14

Figure 2.12 The frequent conditional FP-Tree of “m”. 15

Figure 2.13 Comparing the size of FP-Tree with the order of support of items and the lexicographic. (a) is for support order and (b) is for lexicographic order. 17

Figure 3.1 A database with 5 transactions and the set of all items is I={1, 2, 3, 4, 5}..... 21

Figure 3.2 The algorithm of the Baseline method. 22

Figure 3.3 The pseudo code for counting the support values of candidate supersets..... 23

Figure 3.4 The Baseline method. 24

Figure 3.5 The result of the Baseline method. 25

Figure 3.6 Algorithm Apriori-C. 27

Figure 3.7 An algorithm for counting complement-support. 28

Figure 3.8 An example of Apriori-C. 29

Figure 3.9 The prefix trie when counting the complement-support of C_2 in the example of Figure 3.8. 30

Figure 3.10 An example for Eclat-C. 32

Figure 3.11 A diagram of Data complement technique (DCT). 34

Figure 3.12 An example for DCT with Apriori. 35

Figure 4.1 The execution time in different minimum support threshold for the database (a) D10kT10N20 and (b) D100kT60N100. 41

Figure 4.2 The scalability for six algorithms, Baseline, Apriori-C, DCT-Apriori, Eclat-C, DCT-Eclat, and DCT-FPGrowth in different number of transactions. 44

Figure 4.3 The scalability in different number of transaction when the dataset is large. 45

Figure 4.4 Results of the different number of items (dataset=D10kT5, minsup=60%) 47

Figure 4.5 Results of the different number of items (dataset=D10kT60, minsup=60%) 48

Figure 4.6 Execution time in seconds for the different number of items (N-|T|=60, |D|=10k, minsup=60%). 51

Figure 4.7 Execution time in seconds for the different number of items (N-|T|=40, |D|=10k, minsup=60%) 52

Figure 4.8 Execution time for different average size of transactions (smaller dataset). 54

Figure 4.9 Execution time for different average size of transactions (larger dataset). 55

Figure 4.10 Number of Candidates in each pass (Apriori-C). 56

Figure 4.11 Number of Candidates in each pass (Baseline). 56

Figure 5.1 The areas of Distance Learning 59

Figure 5.2 Learning behavior of 3 students, S_1, S_2 and S_3. 60