CS 43105 Data Mining Techniques

Homework 2
Instructor: Xiang Lian
Due Date: Please refer to the course website

1. Implementing Apriori algorithm (using either C++ or Java). Your program should be able to accept two parameters with input: filename and a minimal support level. For instance, "myapriori filename 15", where "myapriori" is the execution file, and 15 means a frequent itemset has frequency of 15% of the entire transactions in "filename". The file format is as follows: each line corresponds to a transaction (**no transaction id**) and each item in the transaction is separated by a space. Your program should output all the frequent itemsets in the input file with the specified minimal support level. **[70 points]**

a) A detailed Pseudo code including the necessary data structure for implementation [20 points];
b) Source code [40 points], and;
c) The running results (screen captures) of the following input file and minimal support (10%, 20%, 30%, 50%) [10 points].
1 2 3 4
1 4 5 6
2 3 4
1 2 3 4
2 3 6
1 2 4 6
4 5
1 2 3 4 5
3 4 5
1 2 3 6
1 2 3 5
1 4 5
2 3 4 6
1 2 3 4
2 3 4
1 2 4 5 6
3 4 5
1 2 3 4 5 6
3 4 5 6
1 2 3 5

2. Given 4 items, A, B, C, and D, list all possible itemsets in a lattice. **[10 points]**

3. In the following transaction database, if the minimum support is 7, please list all frequent itemsets and their support counts. **[20 points]**

TID	Transactions
100	{A, B, E}
200	{B, D}
300	{A, B, E}
400	{A, C}
500	{B, C}
600	{A, C}
700	{A, B}
800	{A, B, C, E}
900	{A, B, C}
1000	{A, C, E}

Bonus Question [extra 20 points]

In Question 3 above, please answer the following questions:

(a) draw the lattice of all itemsets (with their support counts) [10 points];
(b) Given minimal support 5, list closed and maximal frequent itemset(s) [10 points].

Submission

Submit an electronic copy of your homework solution to the Blackboard.