CMSC5724: Exercise List 11

Consider the mining of association rules on the transactions:

transaction id	items
1	A, B, E
2	A, B, D, E
3	B, C, D, E
4	B, D, E
5	A, B, D
6	B, E
7	A, E

Problem 1. What is the support of the itemset $\{B, D, E\}$?

Problem 2. What is the support and confidence of the association rule $BD \to E$?

Problem 3. Consider the application of the Apriori algorithm to find all the frequent itemsets whose counts are at least 3. Recall that the algorithm scans the transaction list a number of times, where the i-th scan generates the set F_i of all size-i frequent itemsets from a candidate set C_i. Show C_i and F_i for each possible i.

Problem 4. Find all the association rules with support at least 3 and confidence at least 3/4. For your convenience, all the itemsets with support at least 3 are $\{\{A\}, \{B\}, \{D\}, \{E\}, \{A, B\}, \{A, E\}, \{B, D\}, \{B, E\}, \{D, E\}, \{B, D, E\}\}$.

Problem 5. If the universe U (the set of all possible items) has size n, prove:

- the maximum number of distinct association rules is $\sum_{a=1}^{n-1} \sum_{b=1}^{n-a} \binom{n}{a}\binom{n-a}{b}$.
- $\sum_{a=1}^{n-1} \sum_{b=1}^{n-a} \binom{n}{a}\binom{n-a}{b} = \sum_{\ell=2}^{n} \binom{n}{\ell}(2^\ell - 2)$.