CZ4032 Data Analytics and Mining

Association Rule Mining: Tutorial 2

Q1 Explain the following observation for PCY algorithm

a. If a bucket contains a frequent pair, then the bucket is surely frequent
b. However, even without any frequent pair, a bucket can still be frequent

Q2 Given a dataset, minsup threshold, which of the following has the largest number of itemsets? Which has the smallest number of itemsets?

2) Frequent itemsets
3) Maximal frequent itemsets
4) Closed frequent itemsets

Q3 Discuss the impact of the following characteristics of a transaction table on the use of the FP tree to mine frequent itemsets from the table:

(a) Number of unique items in table

(b) Average number of items in a transaction

(c) Number of transactions in table

Q4 A database has four transactions. Let *min_sup* = 60% and *min_conf* = 80%.

TID	Date	Items_bought
T100	20006-01-01	{*K*, *A*, *D*, *B*}
T200	20006-01-01	{*D*, *A*, *C*, *E*, *B*}
T300	20006-01-01	{*C*, *A*, *B*, *E*}
T400	20006-01-01	{*B*, *A*, *D* }

(a) Find all frequent itemsets using FP-growth.

Q5: for sequence pattern mining, answer the following questions

- Can <{a},{b},{c}> merge with <{b},{c},{f}> ?
- Can <{a},{b},{c}> merge with <{b,c},{f}>?
- Can <{a},{b},{c}> merge with <{b},{c,f}>?
- Can <{a,b},{c}> merge with <{b},{c,f}> ?
- Can <{a,b,c}> merge with <{b,c,f}>?
- Can <{a}{b}{a}> merge with <{b}{a}{b}>?
- Can <{b}{a}{b}> merge with <{a}{b}{a}> ?