CSE 40647/60647 Data Science (Spring 2018)
Lecture 22: Frequent Pattern Mining: FP-Growth

Frequent itemset mining

- Level-wise, join-based approach: Apriori (Agrawal & Srikant@VLDB'94)
  - Direct hashing and pruning: DHP (Park, Chen, Yu@SIGMOD'95)

- Vertical data format approach: Eclat (Zaki, Parthasarathy, Ogihara, Li@KDD'97)

- Frequent pattern projection and growth: FPgrowth (Han, Pei, Yin @SIGMOD'00)

Closed itemset mining

- Pattern growth-based approach: CLOSET+ (Wang et al. @KDD'03)

Name (NetID):

Given a transaction database:

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Use FP-Growth to find frequent itemsets if min_sup = 2.

Hint: (FP-Growth)

- Find frequent single items and partition the database based on each such item
- Recursively grow frequent patterns by doing the above for each partitioned database (also called ***conditional database***)
- To facilitate efficient processing, an efficient data structure, FP-tree, can be constructed
- Recursively construct and mine ***(conditional) FP-trees***
- Until the resulting FP-tree is empty, or until it contains only one path — single path will generate all the combinations of its sub-paths, each of which is a frequent pattern

Solution:

Answer: 1-itemsets: A:2, B:3, C:3, E:3; 2-itemsets: AC:2, BC: 2, BE: 3, CE: 2; 3-itemset: BCE: 2.