Q5 Report: Frequent Itemset Mining

Author: YANG Rongfeng

Date: 2020-05-31

1. Algorithm Used

I implement FP-Growth algorithm by the pyfpgrowth package to mine all the frequent itemsets with min_sup = 100. The detailed steps of the FP-Growth algorithm are:

Step 1:	Deduce the ordered frequent items. For items with the same frequency, the order is given by the alphabetical order.
Step 2:	Construct the FP-tree from the above data
Step 3:	From the FP-tree above, construct the FP-conditional tree for each item (or itemset).
Step 4:	Determine the frequent patterns.

Then I mine the closed frequent itemsets and maximal frequent itemsets according to their definitions. I will traversal the frequent itemsets that are mined in Task one to find the itemsets that satisfy the definitions.

- An itemset is maximal frequent if none of its immediate supersets is frequent.
- An itemset is closed if none of its immediate supersets has the same support as the itemset.

2. Time Consuming

The time costs of three task are shown in the following table

	Task 1	**Task 2**	**Task 3**
Running Time	179.21(s)	259.63(s)	132.09(s)

Reference

[1] https://fp-growth.readthedocs.io/en/latest/readme.html#getting-started
[2] https://github.com/evandempsey/fp-growth