# FP-growth based frequent itemset mining

V Rajalakshmi, K Vagdevi
August 26, 2018

## 1 IMPLEMENTATION DETAILS

This report describes the algorithm and the implementation details of the FP-growth based frequent itemset mining.
The implementation starts by reading the input dataset file and storing the all the unique items of the data along with their frequencies in the data structure called *itemsCounter*. Then this data structure is sorted according to the frequencies of the items with the help of the function called *sortDesc*. Then, some of the items with the support count lesser than the minimum support count are pruned off and the frequent items are stored in the *itemsCounter* data structure.

Figure 1.1:

| Item | Frequency |
|---|---|
| a | 8 |
| b | 7 |
| c | 6 |
| d | 6 |
| e | 3 |

Filter items based on minimum support count 4 →

| Item | Frequency |
|---|---|
| a | 8 |
| b | 7 |
| c | 6 |
| d | 6 |

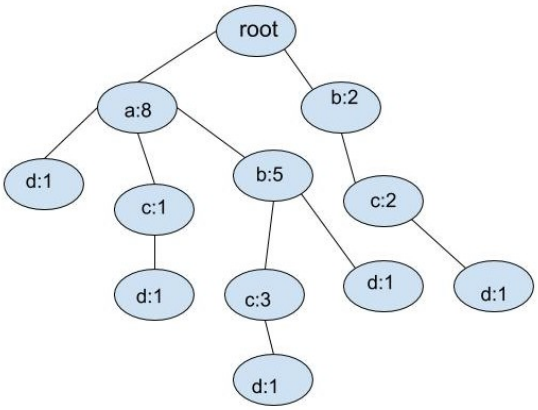
Now, all the transactions are ordered according to the frequencies of the existing frequent items by the function *orderTransaction*, which is implicitly called by the function *orderTable* and these ordered transactions are stored in the data structure called *table*.

Figure 1.2: Ordering transactions

| Transactions | Ordered Transactions |
|---|---|
| a,b | a,b |
| b,c,d | b,c,d |
| a,c,d,e | a,c,d |
| a,d,e | a,d |
| a,b,c | a,b,c |
| a,b,c,d | a,b,c,d |
| a | a |
| a,b,c | a,b,c |
| a,b,d | a,b,d |
| b,c | b,c |

After this, the FP tree of the frequent items is constructed by the function called *constructBaseFPtree*. The construction of the base FP tree provides the basis to mine the tree and extract the frequent item sets recursively.

Figure 1.3: FP tree



Then in the mining part of the FP tree for the frequent item sets, the following happens:

- For each base frequent item, the following steps are followed:
  - The transactions containing the item are extracted from the base FP tree and are stores along with their frequencies.
  - A table containing frequencies of individual items of the transactions are stored and sorted in descending order.
  - The ones with their support counts being greater or equal to the minimum support count are used in the further steps.
  - These items, together with the base item and their combined frequencies are output as part of the frequent item sets.
  - An FP tree is constructed over the new set of frequent items conditioned on the base item.
  - This process is done recursively till root level is reached.
  - Then the frequent item sets are all output along with their frequencies.