Data Mining Project: FP-Growth algorithm Implementation

Yasin Boloorchi

August 17, 2021

1 Introduction

Second section of exercise number 8 request us to implement and test FP-Growth algorithm (just like we did for apriori algorithm in the first section) with the given dataset from http://fimi.uantwerpen.be/data . FP-Growth algorithm is a association rule analysis algorithm (again just like apriori). the main idea aroud FP-Growth is to pruning the dataset graphs; what this algorithm is trying to do, is that at first step, make a tree with the given dataset, after that it will search through nodes with DFS (of bfs, it depends on what programmer wants to use) and save the itemsets that have been seen more frequently in the created tree, and then prune the tree from less frequent itemsets.

2 Dataset

The given dataset is from http://fimi.uantwerpen.be/data , this dataset is contain more than 8 thousand records that can be more than helpfull for testing this implementation of FP-Growth algorithm.

3 Implementation

At the beginning I opened the dataset file, then we have 6 function that we needed. the first function is the one that create the tree, we gave a dataset to it and it will return the tree and all of the items in the datase, next function is the "print-tree" that will print the given tree to it. the "is-it-in" function is a simple function just to check if all of the members of first list are in second list or not. "conditional-fp" function will create dictionary that contains all frequent itemsets and their frequence count. "dfs" function as it's name says, will search the given tree with DFS algorithm. and finally "poss-patt" function will return all possible patterns with the given dataset. after creating tree and search it with DFS, it's time to search all the dataset with the frequent itemset founded and all of their possible patterns and find the hidden rules in the dataset.