Comparison between Apriori and Fp Growth frequent itemset mining datasets.

In this repo I compare the two famous algorithms on the dataset found in this website : http://fimi.ua.ac.be/data/

You can run the script by typing `python3 frequentItemsetMining.py` . The script takes about one hour and then produces a figure like comparison.png included in this repo.

The apriori and fp_growth algorithm implemented are independent. So you can take the any file and use the algorithms like this

```
import fp_growth
minerfp = fp_growth.fpGrowthMiner()
frequentPatterns = minerfp.getFrequentItemset(db,supportThres) #Here db is a list of transaction and supportThres is **percent** support thresold
```

An overview of comparison.

Setting

- I ran the algorithm on 8 different datasets.
- If any of the algorithm takes more than 300 sec on any dataset , I killed it.
- I ran it on different support thresold so I can get to run fp_growth in 0-300 secs.

Observation

- First we can see that apriori runtime grows quadrically with decreasing support thresold. The reason behind this is because with decreasing support , the number of frequent itemsets keep increasing. and the algorithm runtime is dominated by the second phase of the algorithm where we generate all 2-frequent itemsets.
- Secondly we can see that fp_growth runtime doesn't change that much with increasing or decreasing support thresold. Because unlike apriori , fpgrowth has linear lower bound and adds only a constant with varying support thresold.
- Another important observations is the crossing of apriori and fpgrowth. When the support thresold is more , there are much less frequent itemset , in that case apriori beats fp_growth .