Data Mining Project: Apriori algorithm Implementation

Yasin Boloorchi

July 29, 2020

1 Introduction

Exercise number 8 request us to implement and test apriori algorithm. apriori algorithm is a association rule analysis algorithm. the big idea about this algorithm is based on pruning the less frequent itemsets. it's so simple and yet so smart. all we do in this algorithm is that to remove less frequen itemsets from the cycle, and in each iteration of the loop we increase length of itemsets by one. finally after finding all frequent itemsets, we will search all the dataset with possible patterns of this 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 apriori algorithm.

3 Implementation

My whole implementation have been summarized in seven function that will be explained in here. first of all as we know, we shoud open the dataset file. first function if the "get-all-items" that will get all itemsets one by one and store them in a list and return it. ok we have two similar function that will do whole different jobs! one of them is "is-it-in" that will check if all of the members of first list are in second list and the other function that called "is-in-list" will check if the first list is in the second list that been made of lists. "new-candidate-itemset" is the function that generate candidate itemset with the given length. "poss-patt" is a the same function that we will use in the FP-Growth algorithm for making all possible patterns out of given datasets. And finally make-rules function is the one that combine all this function and variables with each others and finally show the founded rules hidden in the dataset.