**Assignment1. Apriori algorithm**

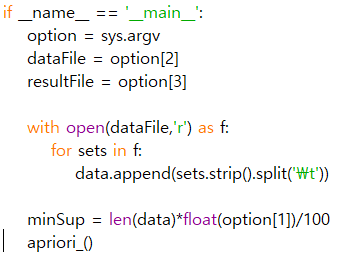
2017029561 Nam Jihoon

1. **Summary**

This algorithm, called apriori algorithm, shows the frequency of occurrence between sets of items. This algorithm derive the frequency of two sets of items common within Transaction, and the probability that sets of conflicting items will appear together when a set appears. This is obtained as Support and Comfort.

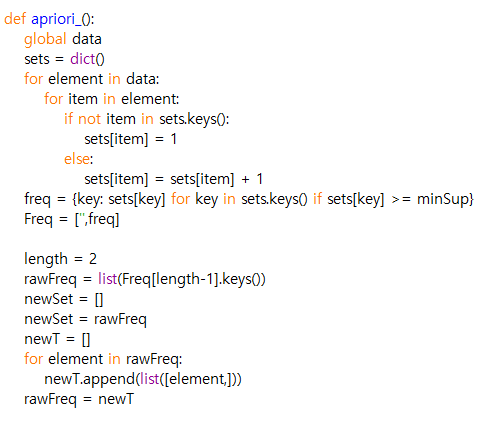
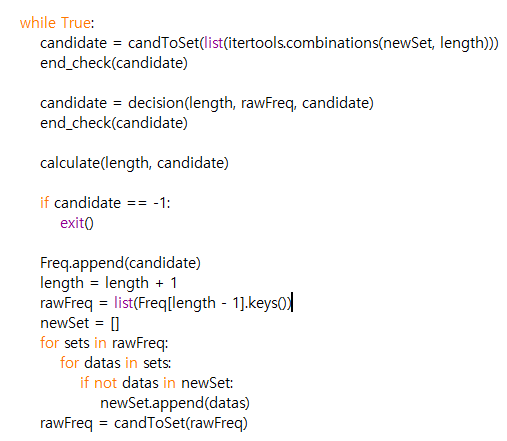
Among them, especially the code I wrote is about a set of items in numbers. For all sets with a minimum support of 5% or more, the Support and Confidence is stored as a text file.

1. **Functions description**
2. **Main function**



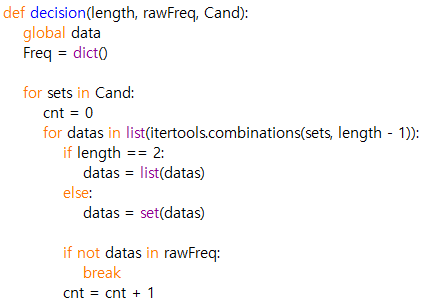
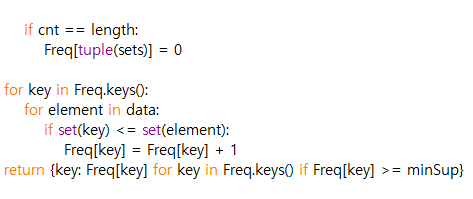
The program is start from main function. If this program is operated with several argument, this function will catch that argument and deal with that. Receives output file name, input file name, and support through command, and gets information from text file through input file name. Then, execute apriori algorhim.

1. **Apriori\_( )**

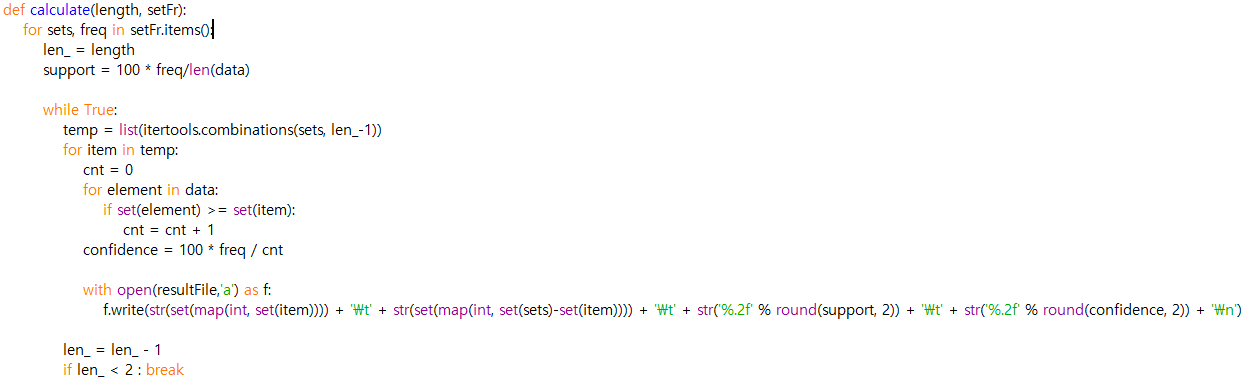
Most progress is made through this function. Use 'dict()' datatype to effectively manage data. If the length is 1, it is different from the normal case, so it is processed before. Then run apriori algorithm until minimum support is not exceeded through the iteration statement. Inside the repeating statement, a set datatype is used for easy processing, based on a set of sets that have often appeared. Then call decision() to prun to select the bandiate. Then calculate the result according to the subset.

1. **Decision( length, rawFreq, Cand)**

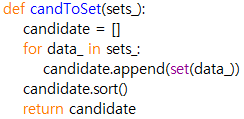
It is one of the functions to find the right one among the candidate. If there are any subset that are not satisfied with the conditions, it is the first way to eliminate them.

1. **Calculate( length, sets )**



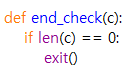
Rotate the subset of the entire item set and compare the relationship between the set of items that are opposite to mine. Calculate support and confidence and save them in textfile.

1. **candToSet( sets\_ )**



Change datatype of candidates to Set

1. **end\_check( c )**



Shut down the program when there is no longer a set of items to run the algorithm.

1. **Compiling Instruction**

The Program written by Python Idle and used Python 3.7.2.

This program was tested on the Windows 10 and executed using command prompt.

[1] download apriori.py and input.txt in same directory

[2] using command prompt, get in to directory that apriori.py is downloaded and enter the command “apriori.py 5 input.txt output.txt”