Apriori Algorithm

2014001312 박지호

1. Summary of the algorithm

This algorithm shows the support rate and confidence rate between the item sets and the associative item sets. Input text file is composed as multiple rows and each row denotes single or multiple item ids. This code implements several proceduress to generate some outcomes.

1. Detailed description of your codes (for each function)

This code has 9 functions including main().

def generate\_subset(given\_set):

This function generates every subsets of given set and returns them except for empty set.

def pre\_process(input\_file):

This function collects the number of item id which exclude duplication.

def scan(trx\_data, item\_id, min\_support):

This function calculates the support rate of item sets and ditch the items lesser than minimum support rate.

def generate\_apriori(lk, k):

This function generates all apriori candidates.

def apriori(item\_id, trx\_data, min\_support):

This function is main frame of apriori algorithm and calls above function as well.

def search\_support(given\_id, final):

This function returns support rate of given item set.

def zip\_item\_and\_support(item\_set, support\_data):

This function links the item set list and support rate list and generate single list.

def save\_file(output\_file, final):

This function literally saves the outcomes as output text file.

def main():

This function receives the parameters from the Terminal.

1. Instructions for compiling your source codes at TA's computer.

Just same as the instruction’s guide. ‘python3 apriori.py 5 input.txt output.txt’. You may change the support rate for sure.

(Python3 version : 3.7.3)

A screenshot of a cell phone

Description automatically generated

1. Any other specification of your implementation and testing

None.