

## 1. 함수 설명

load\_transaction(path, order) 함수

path(input.txt) 파일에 transactions들을 읽어오는 함수이다.

transactions list들과 transactions중에 가장 큰 정수값을 반환한다.(숫자 범위를 알기 위해)

count\_num(itemset, Transactions) 함수

transactions안에 몇개의 itemset이 있는지 즉 얼마나 frequent한 itemset인지 알 수 있는 개수를 반환하는 함수이다.

get\_frequent(itemsets, Transactions, min\_support, prev\_discarded)

k size itemsets들이 Transactions에서 min\_support이상인 itemsets들을 L에 담고 그렇지 않은 k size itemsets들을 new\_discarded로 반환한다.

L에 담긴 itemsets들의 support값 또한 반환한다.

join\_set\_itemsets(set\_of\_itemsets, order)

min support를 만족하는 k-1 size itemsets들을 합집합하여 가능한 k size itemsets 후보 리스트를 만드는 함수이다.

min\_support를 만족하지 않는 k-1 size itemsets들은 맨 끝 요소를 뺀 나머지 요소중에 하나라도 다르다면 frequent하지 않은 후보를 만들게 된다.

이에 따른 join\_two\_itemsets함수에 조건이 주어진다.

join\_two\_itemsets(its1, its2, order)

its1와 its2들중 끝에 한 요소만 다를때 order에 따라 합집합한 결과를 return 그외에는 빈 리스트를 return한다

powerset(s)

어떤 집합 s의 모든 하위집합, 즉 진부분집합을 포함한 부분집합을 애기한다.

itertools에서 combinations, chain을 import하여 구현했다.

write\_rules(X, X\_S, S, supp, conff, num\_trans)

frequent한 X itemsets의 두 부분집합 ( $S+X_S=X$ )으로  $S \rightarrow X_S$  인 association rule의 support와 confidence를 output.txt 포맷에 맞춰서 list로 반환한다.

main()

path = input.txt

min\_support = 5

order = 앞으로 sort할 때 기준이 될 list(처음에는 input파일에 가장 큰 값을 모르므로 max를 100으로 잡았다.)

C = candidate list

L = candidate list중 min\_support 만족하는 요소들 list

supp\_count\_L = L[k] -> k size itemsets들이 transactions에 몇번 나타나는지를 저장하는 list

이 두 list들은 앞에 itemset의 사이즈를 나타내는 index가 있는 list

처음에 C[1]은 0~max값( input파일에 가장 큰 값)이 들어간다.

이중 frequent한 값들을 get\_frequent함수로 f list에 넣고 L[1]에 넣는다.

그리고 supp\_count\_L[1]에 몇번 나타났는지를 저장한다.

그 이후 2이상의 itemset size에 대해서는 itemset\_size를 늘려가며 후보 리스트와 실제 frequent한 값들의 list를 찾는다.

그러다가 L[item\_size]이 없을 때, 즉 item\_size개의 item을 가지는 frequent itemset이 없으면 L[k+1]은 존재할 수 없기 때문에 종료한다.

frequent한 itemsets들을 모두 L리스트에 담았다.

이 중 x가 어느 한 n크기의 itemset이라 할 때

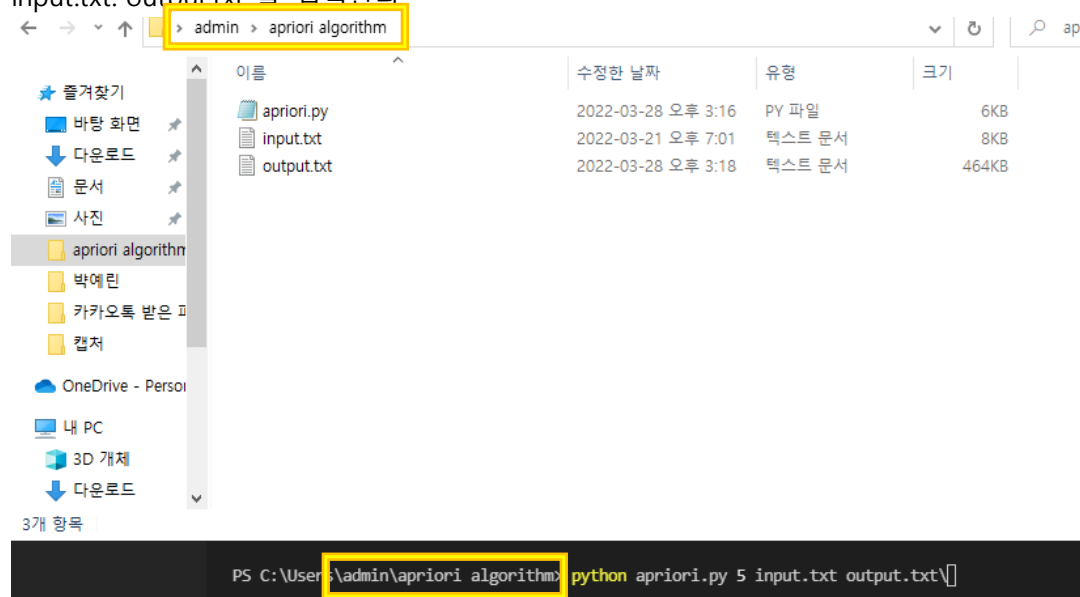
이 itemset의 부분집합들중 한 itemset s가 frequent할 때 x-s itemset또한 frequent한 association rule이 성립할 것이다.

이를 통해 s->x-s의 support값((s + x-s)/x 개수) confidence값(x개수/s개수)를 구해서

write\_rules함수로 output file에 association\_rule string에 넣고 file에 쓴다

## 2. compile방법

Input.txt와 apriori.py을 같은 폴더에 저장하고 그 폴더 directory에서 "python apriori.py 5 input.txt output.txt"를 입력한다



조금 기다리면 해당 디렉토리에 output.txt 가 만들어진다.