Indian Institute of Institute of Information Technology

DMW Assignment 1

Sub dt:19/01/2018

1 Background

The purpose of this programming assignment is to implement the technique of finding frequent itemsets using the *A priori* algorithm and, using this information, generate the association rules which have support and confidence above certain minimum thresholds.

2 Basic assignment

Assume that an input file named `transactions.txt` consists of text that looks as follows:

```
1 3 4
1 2 3 5
2 3 5
2 5
1 2 3 6
```

In the file, blanks separate items (identified by integers) and new lines separate transactions. For example, the above file contains information about a total of 5 transactions and its second transaction consists of 4 items.

Your task is to write a program, in your favorite programming language,[1] that takes as parameters the minimum support, minimum confidence (given as floating point numbers in the range [0..1]), and the name of file of transactions (whose format is as that of the file `transactions.txt` above) and produces *all* association rules which can be mined from the transaction file which satisfy the minimum support and confidence requirements. The rules should be output sorted first by the number of items that they contain (in decreasing order), then by the confidence, and finally by their support (also in decreasing order). An example of a possible session using your program on the data of file `transactions.txt` above is given in Figure 1.

Note: If it makes your life any easier, you can assume that item numbers will be integers in the range $[0..2^{16} - 1]$ and items appear once per transaction and sorted (as above). However, you cannot make any assumptions about the number of transactions that the file may contain.

Students registered in 1DL105 (4pts), shall produce a program which implements the "basic" apriori algorithm and produces the information shown on Figure 1. Students registered in 1DL111 (5pts) will have to do the extra assignment part described in the next section.

```
> myApriori -s 0.25 -c 0.58 transactions.txt
Mined file transactions.txt
and found a total of 16 association rules:
==================================
  Rule      Confidence    Support
==================================
1 2 ==> 3     1             0.4
3 5 ==> 2     1             0.4
1 ==> 2 3     0.666         0.4
1 3 ==> 2     0.666         0.4
2 3 ==> 1     0.666         0.4
5 ==> 2 3     0.666         0.4
2 3 ==> 5     0.666         0.4
2 5 ==> 3     0.666         0.4
1 ==> 3       1             0.6
5 ==> 2       1             0.6
3 ==> 1       0.75          0.6
2 ==> 3       0.75          0.6
3 ==> 2       0.75          0.6
2 ==> 5       0.75          0.6
5 ==> 3       0.666         0.4
1 ==> 2       0.666         0.4
```

There will be plag check for each assignment floated

2. A database has five transactions.Let min.sup=60% and min.conf= 80%.

TID	*Items bought*
T 100	{M,O,N,K,E,Y}
T 200	{D,O,N,K,E,Y }
T 300	{M,A,K,E }
T 400	{M,U,C,K,Y}
T 500	{C,O,O,K,I,E}

i.Find all frequent itemsets using Apriori.