

# Homework 2: Discovery of Frequent Itemsets and Association Rules

## Team

Xiaoxu Gao < [xiaoxu@kth.se](mailto:xiaoxu@kth.se) >

Yu Wang < [wang4@kth.se](mailto:wang4@kth.se) >

## Solutions

- `ReadData` to read transaction data from file "sdata.dat", and processed the transaction line by line then stored it into `trans`.
- `findFrequentItemSets` to find frequent item sets in `trans` which make satisfied minimum support threshold.
  - `findFrequentOneItemSets` to find frequent one item set. Based on frequent 1 item set, find frequent 2 item set,....., find L frequent item set until L-1 frequent item set does not exist.
  - `aprioriGenCandidates` use join and prune to generate L candidate set based on L-1 frequent set.
- `genRule` to generate the association rules found in the data that satisfied the specified support and confidence

## Files

- `Apriori.java` Algorithm to Find frequent itemsets and generate association rules
- `ItemSet.java` Extend `TreeSet` to store the data
- `ReadData.java` Handles data read
- `Rule.java` To generate the association rules
- `sdata.dat` dataset used, with one line per transaction and items seperated with space " "
- `result.txt` result of the given dataset with minimum support 2 and minimum confidence 0.7

(Files should be in the same folder)

## Build and Run

```
>javac Apriori.java
```

```
>java Apriori <min_sup> <min_conf>, such as java Apriori 3 0.6,  
or just java Apriori to use default value.
```

Parameters, i.e., minimum support, minimum confidence and data file, could be changed in `Apriori.java`. And default values are `min_sup = 2`; `min_conf = 0.7`; `data = "sdata.dat"`. It can also be specified at terminal as shown above. Example are shown below:

```
1. wangyu@MacBook-Pro: ~/Dropbox/EIT-DMT-KTH/data mining/homework/src (z...
# wangyu @ MacBook-Pro in ~/Dropbox/EIT-DMT-KTH/data mining/homework/src [11:49:
28]
$ ls
Apriori.java ItemSet.java ReadData.java sdata.dat
# wangyu @ MacBook-Pro in ~/Dropbox/EIT-DMT-KTH/data mining/homework/src [11:49:
58]
$ javac Apriori.java
# wangyu @ MacBook-Pro in ~/Dropbox/EIT-DMT-KTH/data mining/homework/src [11:50:
05]
$ java Apriori 3 0.5
# wangyu @ MacBook-Pro in ~/Dropbox/EIT-DMT-KTH/data mining/homework/src [11:50:
22]
$ java Apriori
# wangyu @ MacBook-Pro in ~/Dropbox/EIT-DMT-KTH/data mining/homework/src [11:50:
31]
$
```

## Result

```
result.txt
Frequent Item Sets:
Frequent 1 Item Sets:
[1], 6
[2], 7
[3], 6
[4], 2
[5], 2
Frequent 2 Item Sets:
[1, 2], 4
[1, 3], 4
[2, 3], 4
[1, 5], 2
[2, 4], 2
[2, 5], 2
Frequent 3 Item Sets:
[1, 2, 3], 2
[1, 2, 5], 2
Association Rules:
[5] -> [1], 1.00
[4] -> [2], 1.00
[5] -> [2], 1.00
[5] -> [1, 2], 1.00
[1, 5] -> [2], 1.00
[2, 5] -> [1], 1.00
```

The Result is stored in file `result.txt`, which includes Frequent Item Sets and Association Rules. Frequent Item Sets are represented like “[2, 5], 2”, item sets in bracket pairs followed by the corresponding support for the set. Association Rules are represented like “[5] -> [1, 2], 1.00”