Data Mining - CS F415

Rule Mining using Apriori Algorithm

Assignment - 1

```
Enter The Value for Minimum Support:
0.45
Enter The value for Minimum Confidence:
0.95
The Time elapsed to find all frequent Item Subsets: 1050 milliseconds
physician-fee-freeze=n  ---> democrat           0.9919028340080972
el-salvador-aid=n  ---> aid-to-nicaraguan-contras=y             0.9807692307692307
el-salvador-aid=n  ---> democrat                0.9615384615384616
el-salvador-aid=n aid-to-nicaraguan-contras=y  ---> democrat            0.9656862745098039
el-salvador-aid=n democrat  ---> aid-to-nicaraguan-contras=y            0.985
physician-fee-freeze=n aid-to-nicaraguan-contras=y  ---> democrat               0.9952606635071091
aid-to-nicaraguan-contras=y democrat  ---> physician-fee-freeze=n               0.9633027522935781
physician-fee-freeze=n education-spending=n  ---> democrat              0.9950495049504949
adoption-of-the-budget-resolution=y physician-fee-freeze=n  ---> democrat               1.0
adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y democrat  ---> physician-fee-freeze=n           0.9753694581280788
adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y  ---> democrat           1.0
The Time elapsed for confidence pruning:  9 milliseconds
The Total Time for generating all rules: 1059 milliseconds
```

Team Members

NAME	ID Number
G V Sandeep	2014A7PS106H
Malla Naga Poojitha Reddy	2014A7PS019H
Snehal Wadhwani	2014A7PS430H

Language Chosen

Java (Development Environment - Eclipse)

Pre-processing

- First step was to treat opinions (stance taken) like “handicapped-infants=n” and “handicapped-infants=y” as different items a candidate has in his basket
- Because of this 17 items (16 stances and 1 class variable) were now transformed to a transaction which can contain at most 17 items in the basket.
- It has to be noted that even though there are now 34 items effectively a candidate can still have only 17 items in his basket. This is because he can take only one stance for a given issue and can belong to only one class i.e either a Democrat or a Republican.
- The situation in which a candidate's stance is mentioned as ‘?’ would be treated as [0,0] for both the opinions
- On the basis of this a row in the data which looks like:

  'n','y','n','y','y','y','n','n','n','n','n',?,'y','y','y','democrat' becomes:

  [1,0,0,1,1,0,0,1,0,1,0,1,1,0,1,0,1,0,1,0,1,0,1,0,0,0,0,1,0,1,0,1,1,0]

- This array of integers is then processed to form a set of integers representing the transaction :

  [0,3,4,7,9,11,12,14,16,18,20,25,27,29,30]

- The entire Apriori Algorithm then works on this representation of the data

Compilation Steps

- javac RuleMining.java
- Java RuleMining

The program will then ask for the minimum Support value and minimum Confidence value. Please a decimal number in the range of 0<num<=1 in this case

Support and confidence value at which interesting rules are generated

Support : 0.45

Confidence : 0.95

1. physician-fee-freeze=n ; ---> democrat 0.9919028340080972
2. el-salvador-aid=n ; ---> aid-to-nicaraguan-contras=y 0.9807692307692307
3. el-salvador-aid=n ; ---> democrat 0.9615384615384616
4. el-salvador-aid=n ; aid-to-nicaraguan-contras=y ; ---> democrat 0.9656862745098039
5. el-salvador-aid=n ; democrat ; ---> aid-to-nicaraguan-contras=y 0.985
6. physician-fee-freeze=n ; aid-to-nicaraguan-contras=y ; ---> democrat 0.9952606635071091
7. aid-to-nicaraguan-contras=y ; democrat ; ---> physician-fee-freeze=n 0.9633027522935781
8. physician-fee-freeze=n ; education-spending=n ; ---> democrat 0.9950495049504949
9. adoption-of-the-budget-resolution=y ; physician-fee-freeze=n ; ---> democrat 1.0
10. adoption-of-the-budget-resolution=y ; aid-to-nicaraguan-contras=y ; democrat ; ---> physician-fee-freeze=n 0.9753694581280788
11. adoption-of-the-budget-resolution=y ; physician-fee-freeze=n ; aid-to-nicaraguan-contras=y ; ---> democrat 1.0

No of Rules Generated

11