House of Representatives Congressional Voting Record

Kevin Kuo

Computer Science Department

Towson University

Towson, Maryland

kkuo1@students.towson.edu

*Abstract*— This report analyzes a data set of 16 critical votes based on the Congressional Quarterly Almanac made by all 435 Representatives from the House of Representatives. We will analyze this dataset to determine any association rules or trends based on how each representative voted.

Keywords— politics, voting, House of Representatives, association

# Introduction

This report studies the voting record of all 435 Representatives from the House of Representatives in the 2nd session of the 98th Congress.

## Dataset

This dataset, “1984 United States Congressional Voting Records Database”, was provided by Jeff Schlimmer on April 27, 1987. There are 17 attributes:

1. Class Name: 2 (democrat, republican)

2. handicapped-infants

3. water-project-cost-sharing

4. adoption-of-the-budget-resolution

5. physician-fee-freeze

6. el-salvador-aid

7. religious-groups-in-schools

8. anti-satellite-test-ban

9. aid-to-nicaraguan-contras

10. mx-missile

11. immigration

12. synfuels-corporation-cutback

13. education-spending

14. superfund-right-to-sue

15. crime

16. duty-free-exports

17. export-administration-act-south-africa

Each of these attributes with the exception of attribute #1 is categorized into either yay, nay, or no response.

## Objective

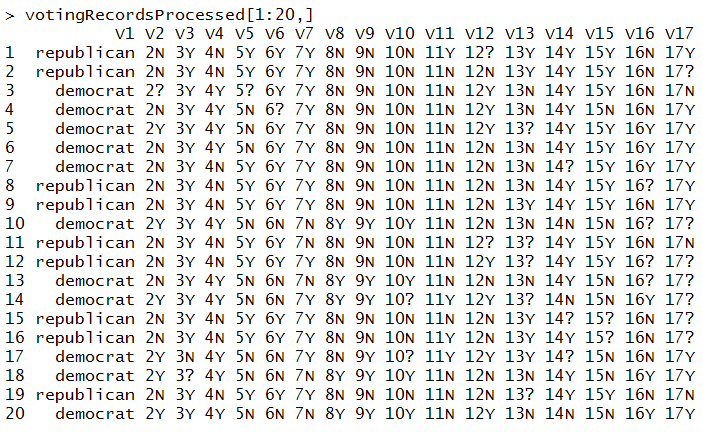
We will analyze this dataset to determine any frequent itemsets and mine for any association rules. This allow us to analyze voting trends.

# Methodology

This section will explain the two association methods used on the dataset and the approach that we have taken to mine the dataset and any pre-processing that was done on the dataset.

## Pre-processing

Minor pre-processing was needed on this data set. The format from the raw data file did not include header labels which makes the data not human readable. In order to make it human readable, it was converted to an Excel spreadsheet format using the ‘xlsx’. Also, each column is a separate vote but without a way to differential values (y, n, or ?) from those in another column we were not able to apply the association algorithms appropriately. I prefixed each value in the column with the column integer label. Also, in order to do association mining, the table must be converted into a transaction data type by first basketing.



## Apriori

We will use the ‘arules’ package in order to apply the apriori pruning principle. This algorithm generates candidate item sets and tests the candidates against original dataset and terminates when no frequent set can be generated – pruning. This association rule minig creates a large number of rules so we use take the subset of this itemset to produce a separate subset of rules. The variable on the right hand side will either be “democrat” or “republican”. The parameters such as support and confidence were determined based on sufficient rules output as well as processing time.

## Eclat

Eclat is part of the ‘arules’ package and it uses the intersection for equivalence class clustering along with a bottom up lattice traversal. The parameters such as support and confidence were determined based on sufficient rules output as well as processing time.

­

# Results

The dataset was a good dataset in which to perform association rules mining/analysis to begin with.

## Item Frequency

The item frequency plots show us which items are important in the dataset.

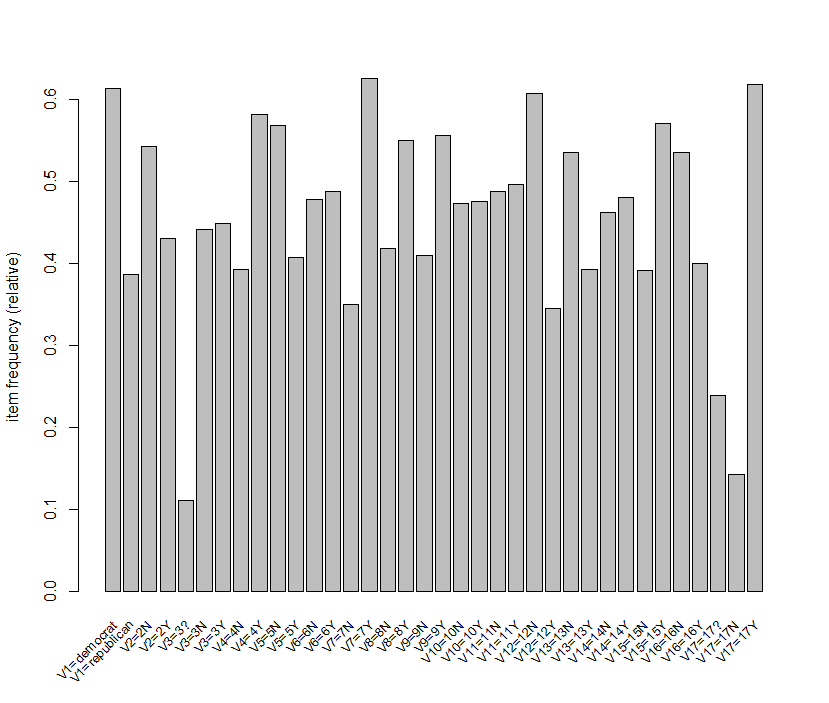


Figure : Item Frequency Plot with Support = 10%

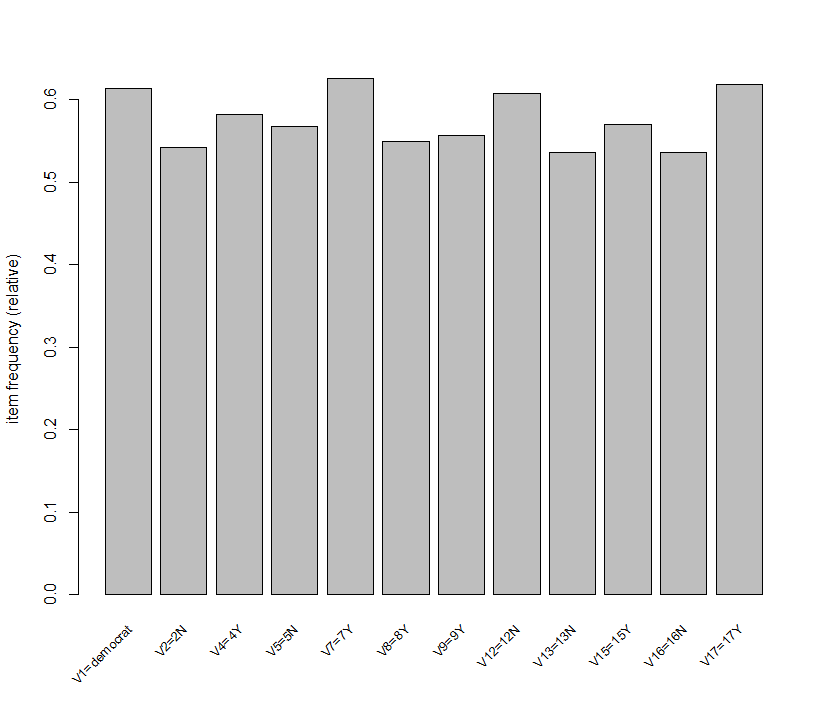


Figure : Item Frequency Plot with Support = 50%

The votes (attributes) that may be important to the dataset using the 50% support parameterare as follows:

2. handicapped-infants: N

4. adoption-of-the-budget-resolution: Y

5. physician-fee-freeze: N

7. religious-groups-in-schools: Y

8. anti-satellite-test-ban: Y

9. aid-to-nicaraguan-contras: Y

12. synfuels-corporation-cutback: N

13. education-spending: N

15. crime: Y

16. duty-free-exports: N

17. export-administration-act-south-africa: Y

## Apriori

We analyzed the general Apriori association rules with the entire itemset.

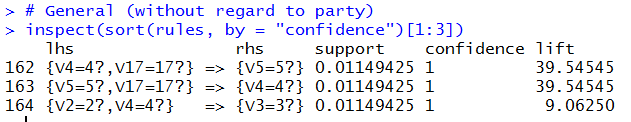


Figure : Apriori association on entire Item Set

The relationship among these abstentions (x?) is interesting.

2. handicapped-infants

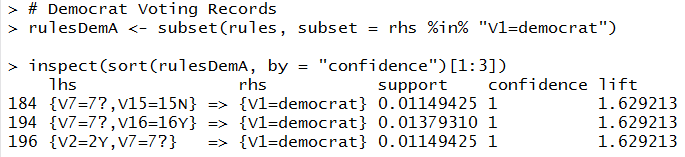
4. adoption-of-the-budget-resolution

5. physician-fee-freeze

17. export-administration-act-south-africa

They appear to be vastly different pieces of legislation and the question would be if the votes for all 4 of these occurred on the same day in which these representative were absent of the legislative session.

### Democrat Subset



The top 3 association rules for this subset appear to be the following if the Congressman is a Democrat:

2. handicapped-infants: Y

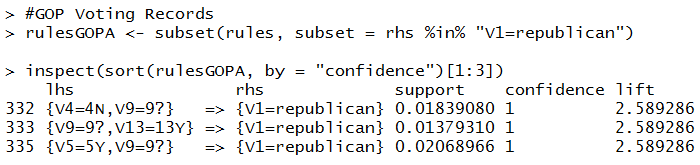
7. religious-groups-in-schools: ?

15. crime: N

16. duty-free-exports: Y

Most of these seem to be consistent with the policies of the Democratic party with the exception of the abstention from religious groups in schools. One wonders whether these abstentions came from Congressmen representing a particular region of the country such as the Southeast.

### Republican Subset



The top 3 association rules for this subset appear to be the following if the Congressman is a Republican:

4. adoption-of-the-budget-resolution: N

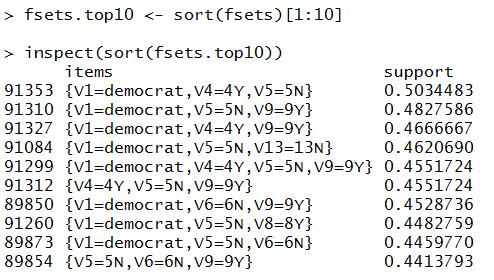
5. physician-fee-freeze: Y

9. aid-to-nicaraguan-contras: ?

13. education-spending: Y

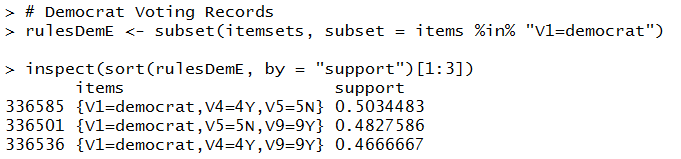
## Eclat

The general Eclat algorithm results are below.



There are a few interesting observations about the Eclat algorithm. First, it seemed to find far more association rules that involved the party attribute being Democrat. This could partly be the result of there being less Democrats in office during that time and them needing to work with broad party support on these key votes. Another interesting observation is that the important votes (attributes) don’t necessarily match up with the top 3 from the Apriori Democrat subset. However, the way in which the Democrats voted in the Eclat general trends seems to be opposite of the typical way Republicans voted in the Apriori Republican subset.

### Eclat Democrat Subset



The top 3 associations in the Eclat algorithm didn’t involve the top 3 attributes from the Apriori Democrat subset which is similar to what we saw above in the general assocations. In fact, the top 3 general Eclat associations are the same as the top 3 in the Eclat Democrat subset.

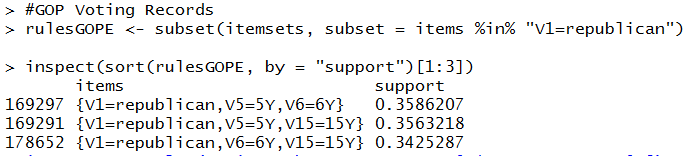
The top 3 association rules for this subset appear to be the following if the Congressman is a Democrat:

4. adoption-of-the-budget-resolution: N

5. physician-fee-freeze: N

9. aid-to-nicaraguan-contras: Y

### Eclat Republican Subset



The top 3 association rules for this subset appear to be the following if the Congressman is a Republican:

5. physician-fee-freeze: Y

6. el-salvador-aid: Y

15. crime: Y

# conclusion

All the association mining algorithms worked well this dataset. The association rules uncovered during this analysis provided many expected results as a few unexpected outcomes. Further analysis into voting records can help historians and researchers identify political trends, anomalies, and environment during these years.