Analysis and Association Rule Mining of 1984 Congressional Voting Data

Robert Brasso

***Abstract*— Congressional voting on bills in the United States determines the majority of laws that are created in our country (the exception being executive orders). Since this is such a fundamental part of our government, analysis of the voting patterns of congress should give us a better understanding of how our laws are created and help inform us on who we would like to represent us. This paper will use a dataset of congressional voting results from 1984 in an attempt to mine association rules and find hidden patterns in congressional voting.**

I. Introduction

One of the most important aspects of United States democracy is that bills are proposed, voted on by the House of Representatives and Senate, and if passed in both, are then signed into law by the President of the United States. With the exception of executive orders, all of the laws follow this process. This emphasizes the magnitude of process of voting on bills by the House of Representatives and the Senate. We elect officials based on the principle belief that they share our ideologies and will vote for laws based on our shared beliefs.

However, how often do they vote for what we believe in? Do they accurately follow the same patterns in voting that we would if we were in their positon? The central purpose of this paper is to attempt to uncover patterns in congressional voting to gain a better understanding into our law-making process. Do politicians vote based solely on party lines? Are there shared voting patterns in party lines? Are there hidden patterns that politicians follow regardless of party lines?

In order to answer these questions, I will analyze a dataset that contains 16 voting results of the House of Representatives from 1984. The dataset will be described in detail and t-weights along with information gain analysis will be used for the exploratory data analysis. We will attempt to mine association rules using the A-Priori algorithm and machine language suite WEKA (Waikato Environment for Knowledge Analysis). Lastly, we will use a decision tree model in conjunction with WEKA to attempt to classify the politicians political party based on their voting patterns.

II. Dataset Description

The dataset for this paper is a collection of voting results on sixteen bills from the 1984 House of Representatives during the 98th Congress, 2nd Session. The first column name is for the political party and the remaining columns contain the voting results for each of the proposed bills. The voting results are coded as either y, n, or ?. Y is for Yea and represents that they either voted for, paired for, or announced for the bill. N is for Nay and represents that they either voted against, paired against, or announced against the bill. Lastly, ? indicates that they voted present, voted present to avoid conflict of interest, or did not vote. Below I will describe each of the column names for the dataset:

- **Party –** Political affiliation of the house member. Either Democrat or Republican.
- **Handicapped_infants – H.R. 808** Handicapped Infants Protection Act of 1982 - Amends the Child Abuse Prevention and Treatment Act to require the National Center on Child Abuse and Neglect to conduct a study of child abuse or neglect in federally assisted or operated health care facilities…
- **Water_project – H.R. 3678** Water Resources Conservation, Development, and Infrastructure Improvement and Rehabilitation Act of 1983 - Imposes a ceiling on amounts authorized for projects under this Act, subject to specified exceptions…
- **Budget_resolution – H.R. 5247** Congressional Budget Act Amendments of 1984 - **Title I: Congressional Budget Process** - Amends the Congressional Budget Act of 1974 to revise the timetable with respect to the congressional budget process to eliminate the second concurrent resolution on the budget…
- **Physician_fee – H.R. 4136** Medicare and Medicaid Budget Reconciliation Amendments of 1983 Title I: Medicare Reconciliation Amendments - Part A: Payment and Coverage - Related Changes - Directs the Secretary of Health and Human Services to establish a national fee schedule for diagnostic laboratory tests for which payment is made under part B (Supplementary Medical Insurance) of title XVIII of the Social Security Act. Directs the Secretary to set the fee schedule at 60 percent of the prevailing charges paid under part B for similar diagnostic laboratory tests during the fee screen year beginning July 1, 1983…
- **Elsalvador_aid – H.R. 1271** Amends the International Security and Development Cooperation Act of 1981 to allow the President to make the fourth certification which is required for continuing aid to El Salvador only if the certification includes a determination by the President that El Salvador has: (1) made good faith efforts since the last certification to investigate and prosecute those responsible for the murders of seven U.S. citizens;

and (2) taken all reasonable steps to investigate the murder of Michael Kline…

- **Religious_schools – H.R. 4996** Religious Speech Protection Act - Prohibits federally funded public secondary schools which allow non-school-sponsored groups of students to meet from discriminating against any meeting of students on the basis of religious content if: (1) the meeting is voluntary and student initiated; (2) there is no government sponsorship; and (3) no unlawful activity is permitted…
- **Anti_satellite – H.R. 5571** Arms Race Moratorium Act - Expresses the sense of the Congress that the President should immediately communicate to the Soviet Union the willingness of the United States to enter into a mutual United States - Soviet Union moratorium on the flight testing and deployment of new ballistic missiles and anti-satellite weapons and the testing of nuclear warheads…
- **Contras_aid – H.R. 2968** Limits the amount that may be obligated or expended for covert assistance for military operations in Nicaragua…
- **Mx_missle – H.R. 2366** Permits the use of specified appropriations to pay the claims of certain Indian tribes for expenses incurred by such tribes for community impact planning activities relating to the potential deployment of the MX missile system…
- **Immigration – H.R. 1510** Immigration Reform and Control Act of 1983 - Title I: Control of Illegal Immigration-Part A: Employment - Amends the Immigration and Nationality Act to make it unlawful for a person or other entity to: (1) hire, or recruit, or refer for a fee for U.S. employment any alien knowing that such person is unauthorized to work, or any person without verifying his or her work status (applies to employers of four or more employees); or (2) continue to hire an alien knowing of such person's unauthorized work status…
- **Synfuels_cutback – H.R. 4098** Synthetic Fuels Corporation Fiscal Accountability Act of 1983 - Amends the Energy Security Act to prohibit the U.S. Synthetic Fuels Corporation from making new awards of financial assistance after the date of the enactment of this Act and before the date on which the Corporation's comprehensive strategy for achieving the national synthetic fuel production goal is approved by Congress…
- **Ed_spending – H.R. 659** National Education and Economic Development Act of 1983 - Provides for Federal assistance for improved elementary and secondary school programs in mathematics, science, technology, and foreign languages…
- **Right_to_sue – H.R. 5640** Superfund Expansion and Protection Act of 1984 - Includes among specified objectives of this Act the creation of a waste end tax on the land disposal of hazardous substances which will discourage the environmentally unsound disposal of hazardous substances and provide additional revenues for the Hazardous Substance Superfund…
- **Crime – H.R. 5690** Anti-Crime Act of 1984 - Title I: Bail - Bail Reform Act of 1984 - Repeals the Bail Reform Act of 1966 and sets forth new bail procedures. Authorizes a judicial officer to consider the safety of any person or the community when making a pretrial release determination…
- **Duty_free – H.R. 2471** Amends the Tariff Schedules of the United States to make duty-free the rendering of geophysical or contracting services in connection with the exploration or extraction of natural resources…

- **Southafrica_export – H.R. 4230** Export Administration Amendments Act of 1984 - Title I: Amendments to Export Administration Act of 1979 - Amends the Export Administration Act of 1979 (the Export Administration Act) to amend the congressional findings and declaration of policy with respect to export controls…

III. Data Preprocessing

The dataset for this project was complete, however it was missing the required headers to make it easier to import using python. I manually added the headers to the csv file. Additionally, there are many values that are defined as ?, but since these don't represent missing data they were left in the dataset.

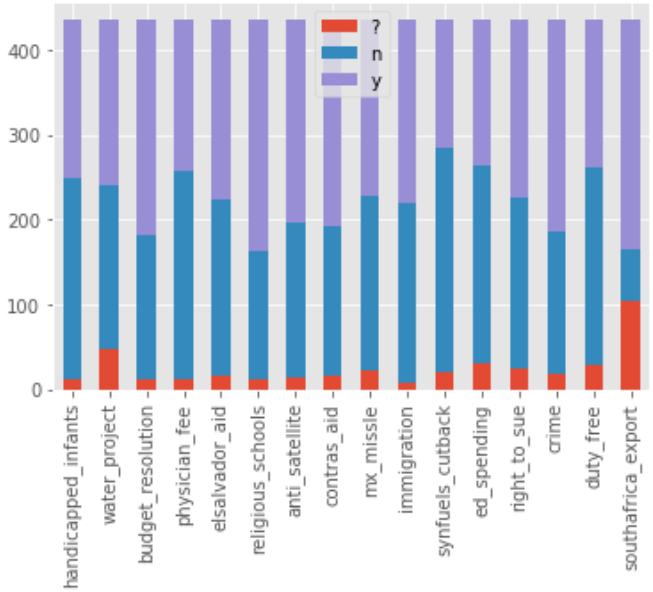
IV. Exploratory Data Analysis

One of the key questions I have about this dataset prior to starting to extract association rules is what the results of the voting was on each of the given topics. In order to explore this I simply took the total counts for each of the votes and created a table, listed below.

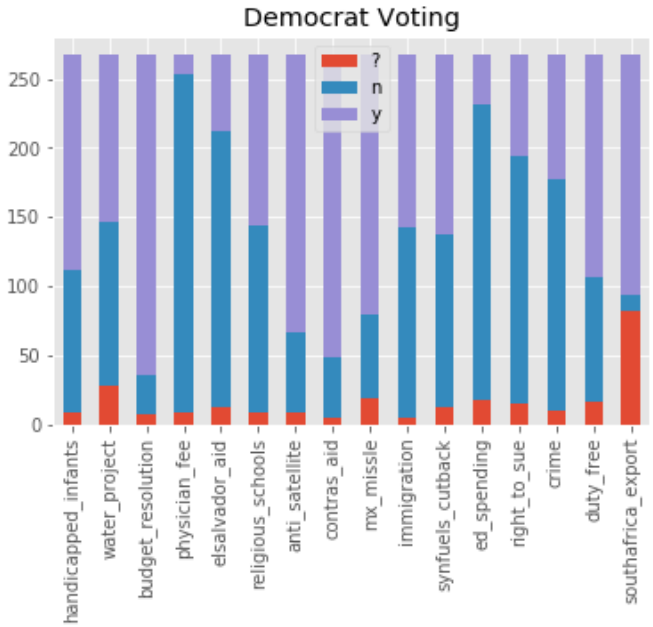
Bill Topic	Y	N	?
Handicapped_Infants	187	**236**	12
Water_project	**195**	192	48
Budget_Resolution	**253**	171	11
Physician_fee	177	**247**	11
Elsalvador_aid	**212**	208	15
Religious_schools	**272**	152	11
Anti_Satellite	**239**	182	14
Contras_aid	**242**	178	15
Mx_missle	**207**	206	22
Immigration	**216**	212	7
Synfuels_cutback	150	**264**	21
Ed_spending	171	**233**	31
Right_to_sue	**209**	201	25
Crime	**248**	170	17
Duty_free	174	**248**	28
Southafrica_export	**269**	62	104

Based on the table, we can see that 11 of the bills were passed and 5 were not. Additionally, we can see that thewater_project, elsalvador_aid, mx_missle, immigration, and right_to_sue bills were very closely contested. We also see that the southafrica_export, water_project, ed_spending, duty_free, and right_to_sue had the highest number of non yea or nay votes. Based on the table above, I created a stacked

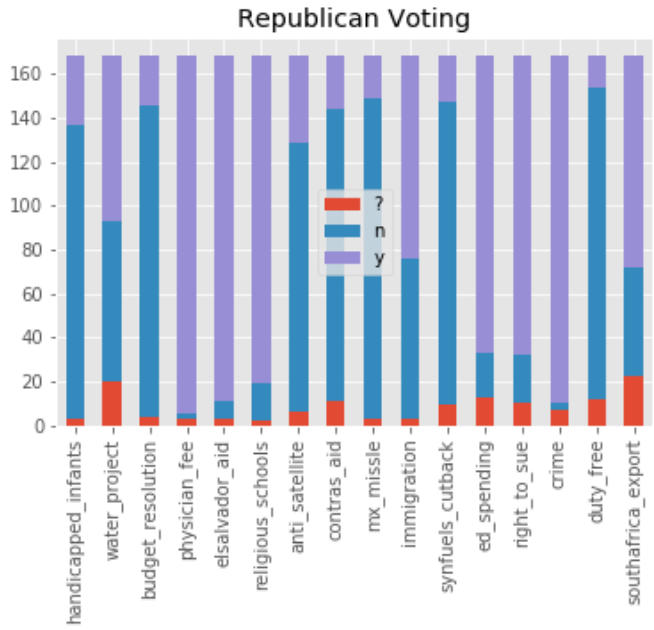
barchart to better visualize the data.



After getting a feel for the overall patterns, I felt it was important to attempt to identify patterns in party voting. To this extent I made stacked barcharts for both democratic and republican voting results.



Looking at the democrat voting, we can see that they voting in favor of handicapped_Infants, water_project, budget_resolution, anti_satellite, contras_aid, mx_missle, synfuels_cutback, duty_free, and southafrica_export. Democrats voted against physician_fee, elsalvador_aid, religious_schools, immigration, ed_spending, right_to_sue, and crime. Synfuels_cutback, immigration, religious_schools, and water_project were the most closely contested votes.



Looking at the republican voting, we can see that they voted in favor of the water_project, physician_fee, elsalvador_aid, religious_schools, ed_spending, right_to_sue, crime, and southafrica_export. They voted against handicapped_infants, budget_resolution, anti_satellite, contras_aid, mx_missle, synfuels_cutback, and duty_free. The water_project was the only closely contested vote.

Based on analysis of the voting patterns for both parties it appears that the non-partisan issues are water_project and southafrica_export. The remaining issues are partisan issues.

V. Association Rule Mining Results

In order to further investigate this data, I wanted to determine if there were any interesting association rules that could be mined from this dataset. In order to do this, I ran the data through the Apriori algorithm and experimented with different supports to find interesting rules. First, I ran the algorithm with a minimum support of 10% and a minimum confidence of 50%. This first test took several minutes as it generated 577,705 total rules. The rules generated from this support threshold contained very high lift scores, but tended to have very specific antecedent requirements. The top 5 rules from this are listed below:

```
    lhs                          rhs                       support confidence     lift count
[1] {handicapped_infants=y,
     physician_fee=n,
     elsalvador_aid=n,
     anti_satellite=y,
     mx_missle=y,
     immigration=n,
     crime=n}                 => {religious_schools=n} 0.1195402   0.962963 2.755848    52
[2] {handicapped_infants=y,
     physician_fee=n,
     elsalvador_aid=n,
     anti_satellite=y,
     contras_aid=y,
     mx_missle=y,
     immigration=n,
     crime=n}                 => {religious_schools=n} 0.1195402   0.962963 2.755848    52
[3] {handicapped_infants=y,
     budget_resolution=y,
     physician_fee=n,
     elsalvador_aid=n,
     anti_satellite=y,
     mx_missle=y,
     immigration=n,
     crime=n}                 => {religious_schools=n} 0.1195402   0.962963 2.755848    52
[4] {Party=democrat,
     handicapped_infants=y,
     physician_fee=n,
     elsalvador_aid=n,
     anti_satellite=y,
     mx_missle=y,
     immigration=n,
     crime=n}                 => {religious_schools=n} 0.1195402   0.962963 2.755848    52
[5] {handicapped_infants=y,
     budget_resolution=y,
     physician_fee=n,
     elsalvador_aid=n,
     anti_satellite=y,
     contras_aid=y,
     mx_missle=y,
     immigration=n,
     crime=n}                 => {religious_schools=n} 0.1195402   0.962963 2.755848    52
```

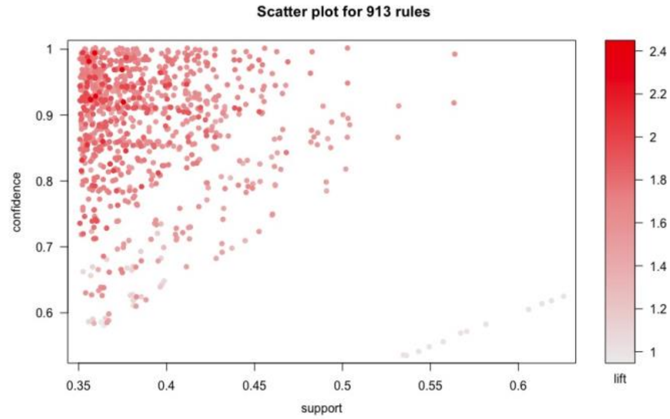
As seen above, the top 5 rules generated all apply towards predicting no votes on the religious_schools bill. This implies that predicting the result of the religious_schools vote based on other voting patterns is easier than predicting the vote of any other bill in the dataset.

Since these rules are somewhat convoluted and only served to predict a no vote on religious_schools, I increased the support threshold a few more times to find an optimum level where the rules were simply, interesting, and contained a high lift score. Through experimentation, the optimum minimum support I arrived at was 35%. Below are the top 10 rules generated:

	lhs		rhs	support	confidence	lift	count
[1]	{Party=republican,elsalvador_aid=y}	=>	{physician_fee=y}	0.3586207	0.9936306	2.441973	156
[2]	{Party=republican,crime=y}	=>	{physician_fee=y}	0.3563218	0.9810127	2.410963	155
[3]	{physician_fee=y,elsalvador_aid=y}	=>	{Party=republican}	0.3586207	0.9285714	2.404337	156
[4]	{physician_fee=y,crime=y}	=>	{Party=republican}	0.3563218	0.9226190	2.388924	155
[5]	{Party=republican}	=>	{physician_fee=y}	0.3747126	0.9702381	2.384483	163
[6]	{physician_fee=y}	=>	{Party=republican}	0.3747126	0.9209040	2.384483	163
[7]	{elsalvador_aid=y,mx_missle=n}	=>	{contras_aid=n}	0.3632184	0.8586957	2.098498	158
[8]	{elsalvador_aid=y,religious_schools=y,crime=y}	=>	{contras_aid=n}	0.3540230	0.8555556	2.090824	154
[9]	{elsalvador_aid=y,crime=y}	=>	{contras_aid=n}	0.3770115	0.8453608	2.065910	164
[10]	{elsalvador_aid=y,mx_missle=n}	=>	{physician_fee=y}	0.3517241	0.8315217	2.043570	153

As shown above, the strongest association rule appears to be that when a congress person is a republican and they voted yes for elsalvador_aid, they were very likely to vote yes on the physician_fee bill. Several of the other rules also support the tie between the physicians_fee bill and republicans. Specifically, the fifth and sixth rules show a direct connection between that bill and the republican party.

In total, by setting the minimum support to 35%, we were able to generate a total of 913 rules. Only the top 18 provided a lift score greater than 2, however there were many rules out of the 913 that showed high levels of confidence and reasonable strong lift scores. A scatterplot of support and confidence, shaded by lift was created to visualize the overall strength of this algorithm when the minimum support threshold is set to 35%.



VI. Conclusion

This paper attempted to identify trends in congressional voting patterns and through analysis of the association rules, there are a few interesting patterns that have emerged. First, it appears that a very large quantity of strong association rules can be pulled from this dataset implying that prediction of congressional voting patterns can realistically provide accurate results. Additionally, it appears that republicans tended to be more predictable in their voting patterns meaning they likely vote among party lines more frequently than democrats. Lastly, it appears significantly easier to predict highly stratified voting results such as religious_schools and physician_fee, and harder to predict highly contested votes like the water_project bill.

References

https://www.congress.gov/bill/98th-congress/house-bill/808

https://www.congress.gov/bill/98th-congress/house-bill/3678

https://www.congress.gov/bill/98th-congress/house-bill/5247

https://www.congress.gov/bill/98th-congress/house-bill/4136

https://www.congress.gov/bill/98th-congress/house-bill/1271

https://www.congress.gov/bill/98th-congress/house-bill/4996

https://www.congress.gov/bill/98th-congress/house-bill/5571

https://www.congress.gov/bill/98th-congress/house-bill/2968

https://www.congress.gov/bill/98th-congress/house-bill/2366

https://www.congress.gov/bill/98th-congress/house-bill/1510

https://www.congress.gov/bill/98th-congress/house-bill/4098

https://www.congress.gov/bill/98th-congress/house-bill/659

https://www.congress.gov/bill/98th-congress/house-bill/5640

https://www.congress.gov/bill/98th-congress/house-bill/5690

https://www.congress.gov/bill/98th-congress/house-bill/2471

https://www.congress.gov/bill/98th-congress/house-bill/4230

Figures