International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

United States Presidential Debates - A Machine Learning and Multivariate Purview

Dr. T. Leo Alexander[1], C. Amelia Betsy[2]

[1]Associate Professor, Department of Statistics, Loyola College, Chennai – 600 034, India

[2]Research Scholar, Department of Statistics, Loyola College, Chennai – 600 034, India

Abstract: ***"Presidential Debate is like a Job Interview, the voters watch out for a trait in the candidates". The entire venture was done using the transcripts of all the candidates.The purpose of this paper is to highlight the significance of the topics spoken and how it affects their victory. Only explicit debates (with the moderator) from 1960 have been taken into consideration.***

Keywords: Candidates, CART- Classification and Regression Tree, Cluster analysis,Principal Components, Text Mining

1. Introduction

Presidential debates are a kind of Super Bowl of American Suffrage. It is so evident that these debates influence the mind of the public. This creates anenigma for the candidates to focus on what they deliver. Various argumentshave impacted the opinion among the public apart from the spontaneity factor. Debates are also meant to be nonpartisan. So there is a great deal of responsibility that lay on the speaker's spotlight. In terms of conviction of the mass public, these play a major role especially for those unaware of the candidates in prior.

Various journals and books were used for reference purpose –"Political Learning from Presidential Debates, Thomas M. Holbrook, Political Behavior Vol. 21, No. 1 (Mar., 1999)", "The Wall Street journal by by Janet Hook and Andrew Ackerman", "Journal of High Technology Management Research 15 (2004) 37 – 50; A text-mining-based patent network: Analytical tool for high-technology trend, Byungun Yoon, Yongtae Park*", "Text Mining: Classification, Clustering, and Applications, edited by Ashok N. Srivastava, Mehran Sahami", "Everitt, B.S. (2005). An R and S-Plus® Companion to Multivariate Analysis. Springer. ciompanion website".

In Sections 2, 3 and 4, various statistical frameworks have been highlighted with respect to Machine Learning and Multivariate Techniques and conclusions are drawn respectively.

2. Frameworks

Three statistical frameworks have been used for analysis purpose. By using Text Mining, all the Presidential transcripts are converted to a data format which is used for subsequent scrutiny.Furthermore, the analysis makes use of CART, Cluster Analysis and PCA.

2.1Text Mining

Text mining is the process of deriving high-quality information from text. It involves the process of structuring the input text, deriving patterns within the structured data, and finally execution and interpretation of the output. The analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics.

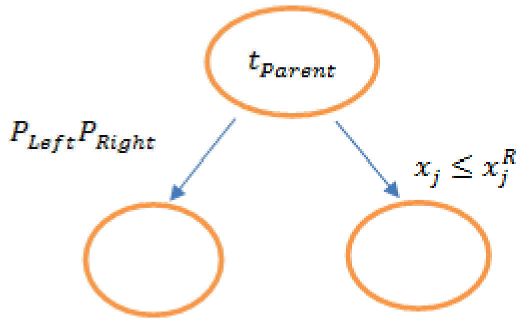
Important tasks in text mining include text categorization, text clustering, concept extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling. The ultimate goal is, to turn text into data for analysis, through application of natural language processing (NLP) and analytical methods.

2.2 Classification and Regression Tree (CART)

Classification and Regression Tree is a classification method which uses historical data to construct so-called decision trees. Decision trees are then used to classify new data. In order to use CART we need to know number of classes a priori. These Decision trees are represented by a set of questions which splits the learning sample into smaller and smaller parts. CART algorithm will search for all possible variables and all possible values in order to find the best split – the question that splits the data into two parts with maximum homogeneity. The process is then repeated for each of the resulting data fragments. CART methodology consists of three parts:

1. Construction of maximum tree
2. Choice of the right tree size
3. Classification of new data using constructed tree.

The splitting rule is given as follows



The Maximization of change of impurity function $\Delta i(t)$ is: $\Delta i(t) = i(t_p) - E[i(t_c)]$, where t_c - left and right child nodes of the parent node t_p. Assuming that the P_l, P_r - probabilities of right and left nodes, we get: $\Delta i(t) = i(t_p) - P_l i(t_l) - P_r i(t_r)$.

2.3 Cluster Analysis

Cluster analysis is a multivariate method which aims to classify a sample of subjects on the basis of a set of measured variables into a number of different groups such that similar subjects are placed in the same group. It is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other clusters. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics. Cluster analysis has no mechanism for differentiating between relevant and irrelevant variables. Therefore the choice of variables included in a cluster analysis must be underpinned by conceptual considerations. This is very important because the clusters formed can be very dependent on the variables included.

In general, if we have p variables $X_1, X_2, \ldots, X_p$ measured on a sample of n subjects, the observed data for subject i can be denoted by $x_{i1}, x_{i2}, \ldots, x_{ip}$ and the observed data for subject j by $x_{j1}, x_{j2}, \ldots, x_{jp}$. The Euclidean distance between these two subjects is given by,

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2}.$$

2.4 Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the smaller of (number of original variables or number of observations). This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables.

The i^{th} Principal Component: Y_i

Select $e_{i1}, e_{i2}, \ldots, e_{ip}$ that maximizes

$$\boldsymbol{var(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = \acute{e}_i \sum e_i}$$

subject to the constraint that the sums of squared coefficients add up to one.along with the additional constraint that this new component will be uncorrelated with all the other defined components.

$$\boldsymbol{\acute{e}_i e_i = \sum\nolimits_{j=1}^{p} e_{ij}^2 = 1}$$

$$\boldsymbol{cov(Y_1, Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{il}\sigma_{kl} = \acute{e}_1 \sum e_i = 0}$$

$$\boldsymbol{cov(Y_2, Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{il}\sigma_{kl} = \acute{e}_2 \sum e_i = 0}$$

$$\vdots$$

$$\boldsymbol{cov(Y_{i-1}, Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{i-1k}e_{il}\sigma_{kl} = \acute{e}_{i-1} \sum e_i = 0}\,.$$

Therefore all the principal components are uncorrelated with one another.

If the variables either have different units of measurement (i.e., pounds, feet, gallons), or if each variable is to receive equal weight in the analysis, then the variables should be standardized before a principal components analysis is carried out. Standardize the variables by subtracting its mean from that variable and dividing it by its standard deviation:

$$\boldsymbol{Z_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j}},$$

where

- X_{ij} $-$ *Data for variable j in sample unit i*
- $\bar{x}_j$ $-$ *Sample mean for variable j*
- s_j $-$ *Sample stanadard deviation for varibale j .*

3. Execution

Conversion of all the transcripts into term document matrix is shown below. This matrix is exclusive of stop words.We are left with 245 terms with only 3% sparsity.

Table 3.1: Term Document Matrix

Term Document Matrix	(terms:245, documents: 26)
Non-/sparse entries	6181/189
Sparsity	3%
Maximal term length	14
Weighting	term frequency (tf)

Looking into the country's issues and various departments, all these terms have been categorised into 5 important variables. The variables State, Defense, Treasury, Commerce, Heath and Human Welfare are the core factors of the country. The rest of the terms have been classified under a general category (Others). The values are given in terms of percentage (standardized form).

Table 3.2 ***Mapping of terms in terms of percentage***

Candidate name	State	Defense	Treasury	Commerce	Health & Human	Others
Kennedy (1960)	17	3	4	2	2	71
Nixon (1960)	13	2	3	1	3	78
Carter (1976)	13	2	5	3	5	71
Ford (1976)	16	2	7	3	3	68
Anderson (1980)	11	1	6	3	2	77
Carter (1980)	15	7	2	2	3	71
Reagen (1980)	10	2	6	2	3	76
Reagen (1984)	7	3	4	2	3	81
Walter (1984)	11	3	3	2	4	77
Bush (1988)	7	3	4	2	3	82
Michael (1988)	10	2	3	2	4	79

Bush (1992)	5	2	5	3	2	83
Clinton (1992)	8	2	6	5	6	73
Peret (1992)	6	1	4	3	4	81
Clinton(1996)	6	2	5	3	5	79
Dole (1996)	7	1	5	3	4	79
Bush (2000)	8	2	5	1	4	79
Gore (2000)	7	3	5	1	3	81
Bush (2004)	7	3	4	2	3	80
Kerry (2004)	8	3	3	2	4	80
Mccain (2008)	10	2	5	3	3	77
Obama (2008)	5	2	4	2	3	84
Obama (2012)	5	2	4	4	3	82
Romney (2012)	7	1	5	5	4	78
Clinton (2016)	7	1	3	3	4	81
Trumph (2016)	6	1	3	2	3	85

To obtain a suitable structure for the candidates (in terms of topic), the table above is comprehended to elucidate the classification and regression tree which is depicted below

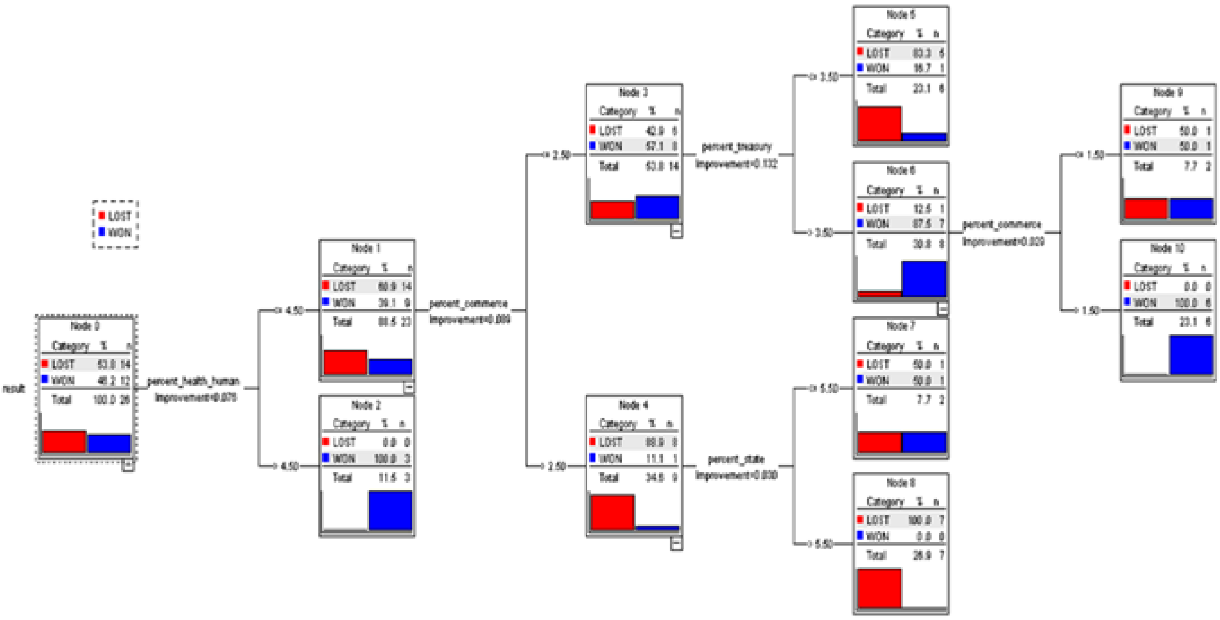


Figure 3.1: Classification and Regression Tree

From the Classification and Regression Tree (CART), we conclude that if a candidate talks less than 4.5% on the topic Health and Welfare and concentrates on Commerce between 1.5% and 2.5% , emphasizing on Treasury more than 3.5%, then there is 100% chance of winning.

If a candidate speaks lesser than 4.5% on health and human welfare, more than 3.5% on Treasury and less than 1.5% on Commerce, then there are only 50% chances to win.

On the other hand, if a candidate talks less than 4.5% on Health and Human Welfare, more than 2.5% on Commerce and more than 5.5% on the State, then he has 100% chances to lose.

For this model, the standard error and its corresponding estimate is tabulated below.

Table 3.3: Standard error and Estimate of the model

Risk	
Estimate	Std. Error
.115	.063
Growing Method: CRT Dependent Variable: result	

We can infer that this model is a good one, as the standard error is at a lower level.

A cross tabulation of the percentage of candidates who have won and lost are given below.

Table 3.4: Cross Tabulation

Classification			
Observed	Predicted		
	Lost	Won	Percent Correct
Lost	12	2	85.7%
Won	1	11	91.7%
Overall Percentage	50.0%	50.0%	88.5%
Growing Method: CRT Dependent Variable: result			

It can be seen that the accuracy of prediction in this model is totally 88.5% which is a positive observation. Proceeding further to cluster the candidates, we have obtained the scree plot. This plot points out to 7 clusters.

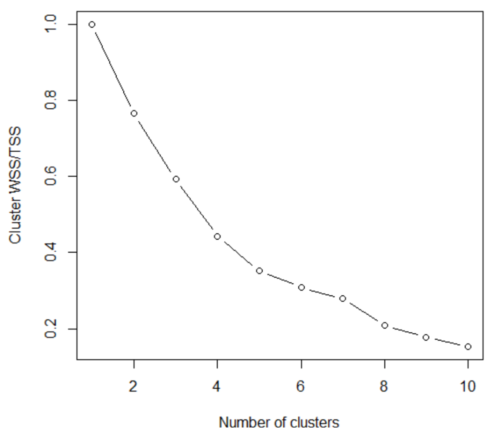


Figure 3.3: Scree Plot

The K-means clustering method highlights the 7 clusters of sizes 2, 3, 9, 1, 4, 2, 5. This is depicted with the aid of Dendogram below.

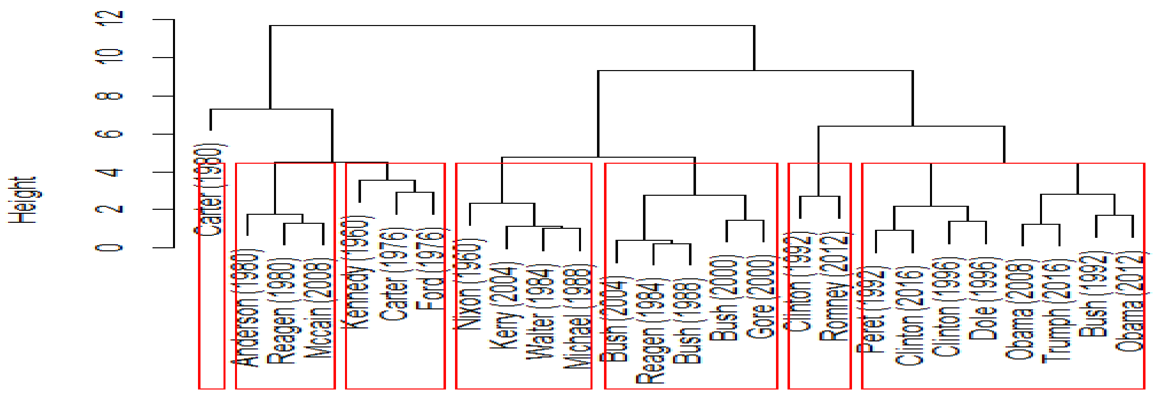


Figure 3.4: Dendogram

As there are a total of 7 clusters, the following table points out to each candidate with their cluster number respectively.

Candidate	Groups
Kennedy (1960)	1
Nixon (1960)	2
Carter (1976)	1
Ford (1976)	1
Anderson (1980)	3
Carter (1980)	4
Reagen (1980)	3
Reagen (1984)	5
Walter (1984)	2
Bush (1988)	5
Michael (1988)	2
Bush (1992)	6
Clinton (1992)	7
Peret (1992)	6
Clinton (1996)	6
Dole (1996)	6
Bush (2000)	5
Gore (2000)	5
Bush (2004)	5
Kerry (2004)	2
Mccain (2008)	3
Obama (2008)	6
Obama (2012)	6
Romney (2012)	7
Clinton (2016)	6
Trumph (2016)	6

In order to show how each candidate is spread over the various topics, the following Biplot(Fig 3.5) is obtained through PCA. From this plot we can interpret how much each candidate has spoken about each topic.

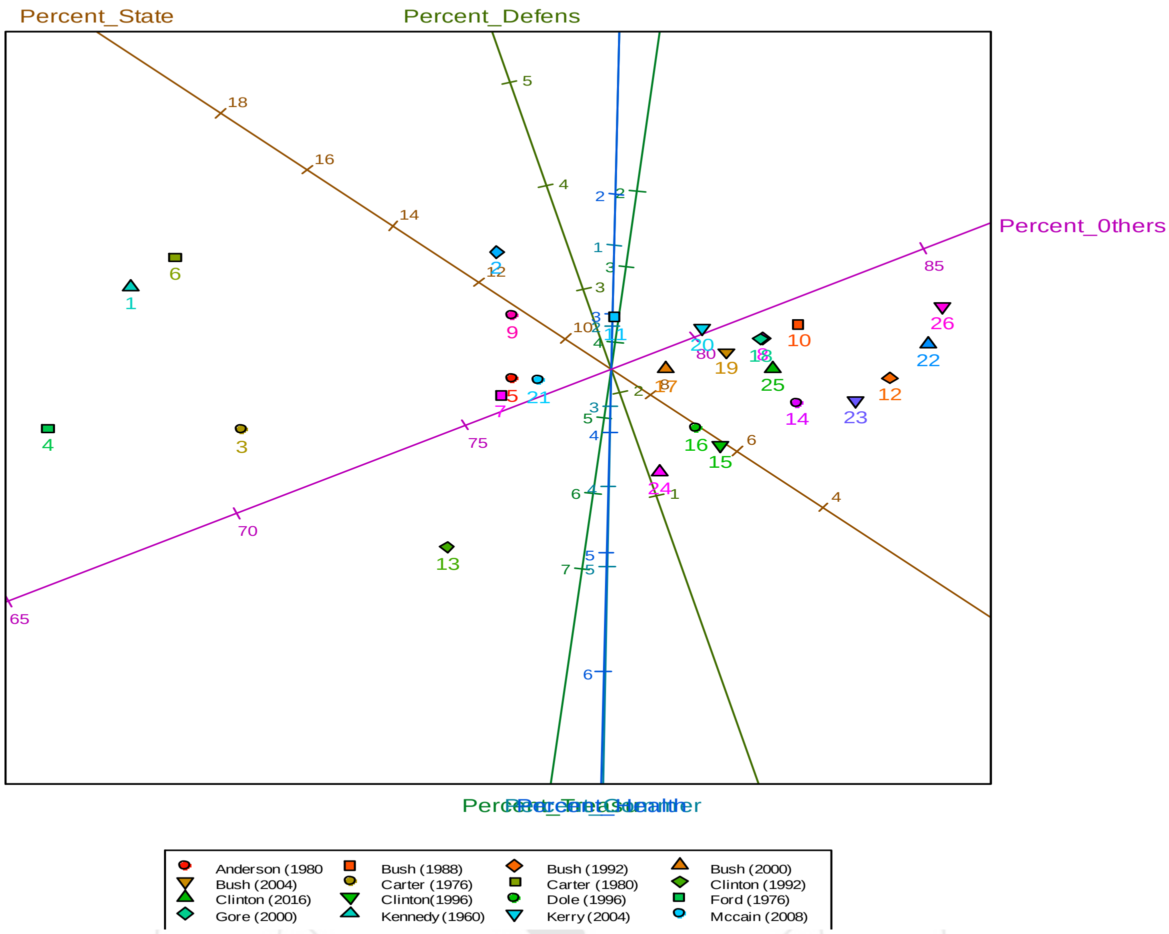


Figure 3.5: Biplot – PCA

If we compare the clusters and the PCA biplot, we can observe that the clustered groups have fallen closer to each other in this biplot. Considering four clusters in random, interpretation (bilpot) for one candidate from each of these clusters is tabulated below.

Table 3.5: Sample coverage of topics

Cluster	Candidate	Coverage of Topics
1	**WINNER** Carter (1976)	Carter has spoken more on State, less on general topic, moderately on Defense and less on the other topics.
2	**WINNER** Nixon (1960)	Nixon concentrates more on Defense, State, Treasury and Commerce.
6	**WINNER** Obama (2008)	Obama speaks more on general topic and very less on the State.
7	**WINNER** Clinton (1992)	Clinton emphasizes more on Commerce, Treasury, moderate on State and less on the other topics.

4. Conclusion

The various topics highlighted above (State, Treasury, Commerce, Defense, Health & Human welfare) falls as the most important ones. Among the 26 candidates right from 1960 to 2016, we have arrived at 6 candidates who seem to have adopted the structure obtained in our analysis (Fig 3.1 Classification and Regression Tree). If the candidates make use of this strategy, then the chances of winning with respect to Presidential Debates is high.

The following table (Table 3.6) shows the list of those 6 candidates spoken above. The weightage given by each of these candidates for all the topics have been tabulated in terms of percentage.

Table 3.6: List of candidates who have followed our structure

Candidate's Name	State	Treasury	Defense	Commerce	Health & Human Welfare	Others	Result
Kennedy (1960)	17%	3%	4%	2%	2%	71%	WON
Reagen (1980)	10%	2%	6%	2%	3%	76%	WON
Reagen (1984)	7%	3%	4%	2%	3%	81%	WON
Bush (1988)	7%	3%	4%	2%	3%	82%	WON
Bush (2004)	7%	3%	4%	2%	3%	80%	WON
Obama (2008)	5%	2%	4%	2%	3%	84%	WON

References

[1] Everitt, B.S. (2005). An R and S-Plus® Companion to Multivariate Analysis. Springer. ciompanion website.
[2] Journal of High Technology Management Research 15 (2004) 37 – 50; A text-mining-based patent network: Analytical tool for high-technology trend, Byungun Yoon, Yongtae Park*
[3] Political Learning from Presidential Debates,Thomas M. Holbrook,Political Behavior Vol. 21, No. 1 (Mar., 1999).
[4] The Wall Street journal by by Janet Hook and Andrew Ackerman.
[5] Text Mining: Classification, Clustering, and Applications, edited by Ashok N. Srivastava, Mehran Sahami.