The Effects of the State of Union Address of USA Presidents to Their Performance Ratings

Carl Lee, Central Michigan University, MI, USA

ABSTRACT

In this article, the author conducts a case study using text mining technique to analyze the patterns of the president's State of the Union Address in USA, and investigates the effects of these speech patterns on their performance rating in the following year. The speeches analyzed include the recent four USA presidents, Bush (1989 – 1992), Clinton (1993 - 2000), G.W. Bush (2001 – 2008), and Obama (2009 – 2011). The patterns found are further integrated and merged with over 4000 surveys on the presidents' performance ratings from 1989 to 2011. Two text mining methodologies are applied to study the text patterns. Two predictive modeling techniques are applied to study the effects of these found patterns to their presidential approval ratings. The results indicate that the speech patterns found are highly associated with their approval rates.

Keywords: Cluster, Cross Validation, Decision Tree, Regression, Text Mining

INTRODUCTION

Text data is a major type of data in the modern society. According to Grimes (2011), over 80% of the data are unstructured, and most of them are text. Analysis of text data started in the 1980’s. Technology advancement in the recent decades has enabled the development of text mining techniques to analyze patterns in text data. Many software, both commercial (e.g., SAS/Enterprise Miner, IBM SPSS, StatSoft) and open sources (e.g., Rapid Miner, Apache UIMA, TESLA), are now available for researchers to analyze text information and further to combine the information retrieved from text data with quantitative data for advanced research. Due to the broad range of applications of text mining technology, Feldman (2004) suggested the use of text analytics in place of text mining. Examples of some common applications include biomedical, linguistic, online text mining, marketing, sentiment, academic, security applications, and so on.

The USA Constitution mandates that the President shall deliver the annual State of the Union Address to the Congress. Today, the State of the Union Address is delivered during the month of January. The Congress Research Service Report for Congress (2006) of USA states
.... the State of the Union Message is a communication from the President of the United States to Congress and the nation in which the chief executive reports on conditions in the United States and, sometimes, around the world, recommends a legislative program for the coming session of Congress, and frequently presents his views about and vision for the present and future.

These annual addresses often have very significant impact to USA as well as to the world since they reflect the current situation of USA and set the direction and policies that the president recommends for the coming year. An interesting question is “Does it matter what the president says in the State of the Union Address?”. In order to answer this question, we first conduct a text mining study to investigate if there are common patterns from these addresses, for example, ‘Are there commonality for the presidents from the same party?’; ‘Are there different patterns between Democrat and Republican presidents?’; and so on. Secondly, we integrate these text patterns with their presidential approval rating data to study if these text patterns have any effects on their performance ratings. Each year, there are a list of surveys regarding the president approval rates in a variety of issues, such as economic, social and health care, budget and debts, foreign affairs, and education issues, etc.. By integrating the patterns found from their State of the Union Addresses from previous year with the performance approval rating survey data, it allows us to study the effects of their addresses to their approval rating on different issues.

The annual State of the Union Addresses that we analyze in this study are from four USA presidents, George Bush (1989 – 1993), William J. Clinton (1993 – 2001), George W. Bush (2001 – 2009) and Barack Obama (2009 – 2011) for a total of twenty three State of the Union Addresses. The documents are downloaded from the American Presidency Project by Woolley and Peters (2011). The approval rating survey data are from the U.S. Officials’ Job Approval Ratings (JARs) project funded in part by a National Science Foundation grant (Niemi, Beyle & Sigelman, 2010). The detailed results of the speech patterns of these 23 addresses have been reported in Lee and Crockett (2012). This article first reviews some of the results from the text pattern analysis and then presents the findings of the effects of these text patterns on the presidential approval ratings. The article is organized as follows. Section 2 briefly describes the speech data preparation and how we integrate the speech data and the approval rating survey data. Section 3 reviews some important patterns found from the text analysis given in Lee and Crockett (2012) that will be used in this effect study. Section 4 describes in detail the manipulation of the approval survey data. Section 5 discusses the methodology of analyzing the effects of the text patterns to the approval rating. Section 6 presents the results and discussions of the text effects.

DATA PREPARATION AND MANIPULATION

Similar to most data mining projects, considerable time is spent in preparing the data for analysis. The raw text files of the State of the Union Addresses are extracted from the project web site maintained by Woolley and Peters (2011). The text mining software used for our analysis is the SAS Enterprise Miner 6.2 (SAS/EM). The first task in the process of data preparation is to create a proper SAS data from the speech documents. The SAS %TMFILTER macro is applied to gather the text file, which is transformed into a specific format accessible by SAS/EM Text Mining technology. Using %TMFILTER macro and SAS/EM, we create a data source that consists of two key variables needed for text mining analysis and for integrating the text data with approval rating survey data. These are ‘FILTERED’ and ‘NAME’ variables. The variable ‘FILTERED’ is linked to where the text data is located, which are in.htm format. The variable ‘NAME’ is the name of the text file that will be needed as the ‘key’
Reduction of Defect Misclassification of Electronic Board Using Multiple SVM Classifiers
www.igi-global.com/article/reduction-of-defect-misclassification-of-electronic-board-using-multiple-svm-classifiers/111448?camid=4v1a

Composition of Optimal Service Workflows with Quality-of-Service Enabled Multi-Criteria Uniform Cost Search Algorithm