Political Sentiment Visualization

Data Analysis and Visualization using Voxgov US Federal Government Media Releases

Yuqi Bi
Columbia University QMSS
517 W113 St. Apt 35, NY, 10025
+1 (517) 691 2154
yb2309@columbia.edu

Darrick Leow
Columbia University QMSS
155 Claremont Ave Apt 630, NY, 10027
+ (347) 703 2135
dyl2123@columbia.edu

Mengying Li
Columbia University QMSS
100 La Salle St. Apt 13C, NY, 10025
+1 (917) 826 1382
ml3576@columbia.edu

Rongyao Huang
Columbia University QMSS
248 W102 St. Apt 4B, NY, 10025
+1 (917) 963 1450
rh2648@columbia.edu

1. INTRODUCTION

Sentiment analyses in politics have only been a recent occurrence in the 21st century. Its development stems from initial text analyses involving product and movie reviews. With time and the rise of social media usage, such as Twitter and Facebook, political science researchers have gradually come to utilize tweets as a viable data source for extracting and conducting sentiment analyses in order to predict or analyze outcomes of various political events, such as opinion polls and election.

With access to a large database of Voxgov federal government media releases from the year 2012, this research paper aims to utilize methods of Structural Topic Modeling (STM) and lexicon-based sentiment extraction in an attempt to visualize potentially interesting political sentiment patterns at a metadata level. The corpus constructed using these federal government media releases provide new insights into political sentiment analysis since it provides a more formal and relevant source of information as compared to the more informal Twitter data used in prior research.

This paper discusses the process in which the aforementioned political sentiment analysis using media releases was conducted. The first section presents a literature review of the prior and existing research related to topic modeling and sentiment analysis in the realm of political science. The second section presents the main motivation for this research paper. This is supported by the forth section, which highlights the main research questions that the paper seeks to address explicitly. The fifth section discusses the dataset and data treatment process in detail. The sixth section provides the methodological steps taken in the data analysis procedure. The sixth section combines the discussion of visualization design implementation, evaluation and presentation of results. The seventh section discusses possible limitations observed in the analysis and visualization process, as well as any future steps for improvement. The eighth section finally concludes this research paper.

2. LITERATURE REVIEW

The literature review may be separated into two main categories, firstly in terms of previous work relating to topic modeling and secondly relating to prior and existing sentiment analysis. It also highlights the potential value-added component of this research paper, namely the utilization of a more formal corpus source (made up from a pool of Voxgov federal government media releases) for political sentiment analysis, as compared to existing research studies that are primarily employ less formal social media (primarily Twitter) sources.

2.1 Generative Topic Modeling

In probability and statistics, a generative model is a model for randomly generating observable data, typically given some hidden parameters. This idea has multiple applications in the field of natural language processing, in particular, topic modeling (Shannon, 1948). The most basic topic model assumes that each document is a mixture of topics and that each word's creation is attributable to one of the documents' topics (Steyvers & Griffiths, 2007). Built on the basic form, Latent Dirichlet Allocation (LDA) allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar (Blei & Jordan, 2002).

A recent advancement in topic models further enables us to associate the document-level meta-data with the distribution of topics across documents and the distribution of words conditioning on a given topic (Roberts, Stewart & Airoldi, 2013). This model, also known as the STM, was chosen for use in this research.

2.2 Political Sentiment Analysis

Early sentiment analyses have revolved mainly around product (Turney, 2002) and movie (Pang et al., 2002) reviews. With the proliferation of social media usage in the 21st century, social researchers have subsequently extended their research on the usage of positive and negative emoticons, as well as hash tags in

Twitter tweets, in aim to extract and determine text sentiment (Go et al., 2009; Pak and Paroubek, 2010; Davidov et al., 2010; Bora, 2012).

Text sentiment analyses have thereafter been extended into the political realm. Over the past years, strong interests have developed further in this area, with a vast number of political science researchers using sentiment analysis in hope to predict election outcomes. An example would be the use of Twitter data in sentiment analysis in hope to predict the 2009 German federal election outcomes (Tumasjan et al., 2010). Using a text analysis software Linguistic Inquiry and Word Count (LIWC) 2007 (Pennebaker et al., 2007), the extraction of sentiment was conducted for over one hundred thousand tweets that were collected between August and September 2009 and contained the names of the six political parties represented in the German parliament. In this research, the authors concluded that the number of tweets mentioning a specific party was directly proportional to the probability of that party winning the election.

Following which, another research by O'Connor et al. (2010) aimed to determine the correlation between public opinion poll outcomes and political sentiment in tweets using a Subjectivity Lexicon (Wilson et al., 2005). Here, the method used to determine sentiment for each day was to allocate a positive score for every positive word, and likewise a negative score for every negative word found within the day's tweets. A sentiment score for each day was calculated by taking the ratio of positive count over negative count thereafter. The research noted that the correlation between twitter sentiment and public opinion polls were strongly correlated on the topic of presidential job approval.

A last example of sentiment analysis involvement in the realm of political research lies within the quest of a few political scientists to develop an accurate classifier to score sentiments more accurately in tweets (Bakliwal et al., 2013). The researchers involved in this study used supervised learning and a combined set of subjectivity-lexicon-based scores, Twitter-specific features and the top 1,000 most discriminative words to develop a classifier which proved to be far superior over other naïve, unsupervised methods of sentiment scoring involving subjectivity lexicons. Such studies highlight the cutting-edge research and continual efforts being channeled into political sentiment analysis, revealing the immense and viable potential for future research development in this subject topic.

Given that most prior sentiment analysis research on text data related to politics have largely utilized Twitter data only, this research paper is the first of its kind in utilizing a large compilation of formal, federal government sourced media text releases in order extract and conduct sentiment analysis on. As such, this contributes largely to the rapidly developing pool of research involving text mining and analysis, while also bridging the gap between data visualization and political science.

3. MOTIVATION

The main motivation for this research is to firstly utilize topic modeling to classify a large, formal text corpus comprising US federal government media releases in the year 2012 into broad topic categories. Secondly, the research intends to thereafter, use text-mining methods to identify and conduct sentiment scoring on each text document within each topic category. This will allow for sentiment analyses to be conducted along various paradigms of metadata components – such as sentiment analyses of media releases across time, location, political party and gender of respondent.

Lastly, the final motivation of this research aims and seeks to present tangible deliverables of the research findings in the form of clear, consistent and easily interpretable data visualizations, produced using various statistical and designing software programs. This will hopefully contribute to the development and advancement in management, analyses and presentation of findings using big data, in the realms of the social and political sciences.

4. RESEARCH QUESTIONS

The first question we aim to solve is what were the main topics potentially discussed in the 2012 Federal Government Release documents?

Following which, the research seeks to answer the question: how does the sentiment, at the level of these broad topical categories, vary across (i) frequency, (ii) time, (iii) political affiliation, (iv) gender of respondent, and (vi) geographic location?

These various research questions will each be addressed separately, with the use different data visualizations techniques in order to present our findings.

5.1 Dataset

Our research utilized the 2012 US federal government daily media releases provided by Voxgov. The original data set contains 200,539 text files in JSON format, which is a hierarchical way of storing data. Each news release will have several {name: value} pairs, including id, date, description (news header), keywords, names, places, mediaTypes, source, and text. The information of interest is stored in the text and source object: text is in html format with tags while the stakeholder information we are interested in, party, gender, location, etc, is listed as sub-attributes within the source feature.

5.2 Treatment of Data

In order to extract the key information required from each news media release – plain text content and the associated party, gender respondent and location details of the source, we removed the html markup in order to separate party, gender and location attributes from other redundant tags in the source object. We accomplished the task using the statistical software R (packages: stringR, Rjson, data.table, etc).

A series of data cleaning and manipulation creates a dataframe comprising 200,539 rows and 15 variables. Since the focus of this research lies mainly in sentiment analysis of news releases and how this correlates with party, gender and location of their sources, we take a subset of the data that has complete information of political party, location and gender attributes. This reduced our dataset size from the original 200,539 to the final 67,678 news release records.

6. METHODOLOGY

Two kinds of modeling techniques are needed in order to answer our research question. They are: (1) structural topic modeling to reveal latent topics of the government news release and classify each document into a topic; and (2) sentiment analysis to evaluate sentiment of each document, each topic, according to their gender, political party or location attribute.

These techniques are discussed in the following sections.

6.1 Structural Topic Modelling

Our general approach is to assume that each news release in the political corpus contains several latent topics, and each topic corresponds to a different categorical distribution of vocabulary. The underlying data generating process can be described as follows:

(1) for each document *i*, draw its distribution of topics Θ_i depending on some parameters;

(2) for each topic *k*, draw its distribution of words Φ_k depending on some parameters;

(3) for each word *n*, draw its topic t_n based on $\Theta_{i;;}$

(4) for each word *n*, draw its realization w_n based on Φ_{tn}.

The difference in the widely used LDA and the newly invented STM approaches lies in how Θ and Φ are determined. LDA assumes that $\Theta \sim \mathit{Dirichlet}(\alpha)$ and $\Phi \sim \mathit{Dirichlet}(\alpha)$, where α and β are fitted with the model. While for STM, the prior distributions for Θ and Φ depend on document-level covariates.

In detail, STM specifies two design matrices of covariates – one for topic prevalence (denoted X), the other for topical content (denoted Z) – where each row defines a vector of covariates for a given document. The topic prevalence component allows the expected document-topic proportions to vary by covariates X rather than arising from a single shared prior. For topic content, it uses Z instead. The following figures 1a and 1b compare the LDA and STM with using a two-plate diagram format. The advantage of STM is clearly depicted.

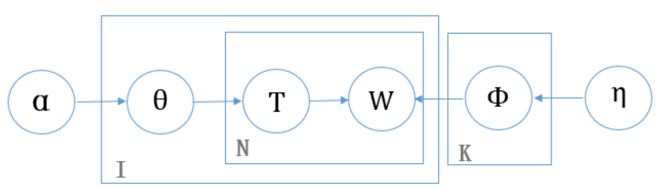

Figure 1a: LDA

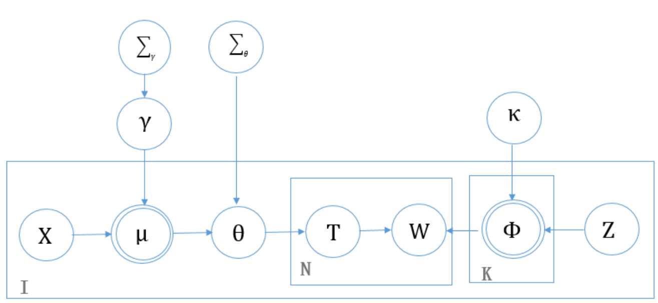

Figure 1b: STM

We implement the STM model with the STM package in R, which gives us results of the topic proportions of each document. The prevalence meta-data used include party, gender and location. For simplicity of analysis, we assign each document to the topic that has the highest proportion.

6.2 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a technique that aims to extract or determine the attitude of a speaker, a writer or a certain subject with regard to some topic or the contextual polarity of a document.

Several existing methods are able to perform automated sentiment analysis of digit texts, among which the *bag-of-words* approach is the most widely used because of its simplicity and efficiency. In this *bag-of-words* model, a text is represented as a multi-set of its words, disregarding grammar and word order. Each of this word is then matched with a sentiment corpus that has a whole dictionary of words labeled as either positive or negative.

In this research, we create our unique political sentiment corpus by combining three dictionaries, namely, Multi-perspective Question Answering (MPQA) Subjectivity Lexicon (Wilson, T., Wiebe, J., and Hoffmann, P.,2005), Bing Liu's Sentiment Lexicon (Liu, B., 2010), and Loughran and McDonald Financial Sentiment Lexicon (Loughran, T. and McDonald, B., 2011).

Based on the matching results, we calculate the raw sentiment score for each document by taking the difference between the number of positive words and the number of negative words and dividing it by the total number of matched words. We then aggregate the document-level sentiment score to topic level using a weighted average. Finally, we normalize the sentiment scores to make it more comparable across topic categories.

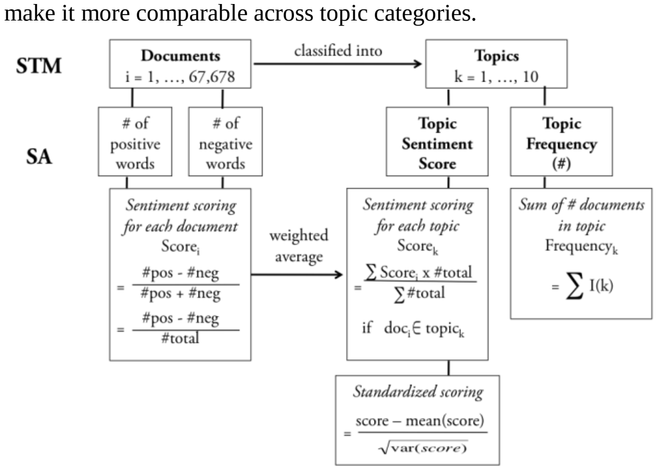

Figure 2: Sentiment Analysis (Sentiment Scoring)

7. VISUALIZATIONS AND RESULTS

7.1 STM Modeling of Topics

Motivation:

This visualization aims to analyze the relationships between clusters of text documents in our political corpus, and to provide a formal clustering of these documents. This would enable us to conduct a "best-guess" naming of each topic category to determine the topical partitioning used in subsequent data visualizations.

Originality of Design:

LDAvis, an R package developed by Sievert, C. and Shirley, K. (2014), is primarily used for visualizing the results from LDA modeling by integrating the visualization application Shiny with D3.js. In order to utilize and customize this package, we converted the previous log-likelihood document-term matrix into percentage form and predicted the most likely topic for each document, using its topic distribution. This enabled us to categorize and sort each text document into their "most-suitable" topic category.

Figure 3a: STM Modeling of Documents into Topics

Aesthetics:

From Figure 3a, we can see that this interactive visualization assigns each topic as a bubble, the size of which reflects the number of documents appearing in that given topic. The positions of these bubbles are based on their similarity with each other based on a specific algorithm. Mouse over each bubble, the relevant 'most frequent'[1] words within that specific topic will be presented on the right side of the canvas. Analyzing the words that appear in each topic enables us to use a "best guess" method of determining the most appropriate label for each topic category.

There exist various options to customize the visualization by exploring different paradigms using filters located at the top of the application layout. The first three filters enable users to analyze the relationship between all topics by using different algorithms. This would cluster the topic categories differently in terms of distances between clusters and grouping of clusters. The last two filters are designed to control the outcome layout of the words within each topic category, i.e. the number of words to be presented on the right side, and the how unique these words are within each topic, relative to the whole political corpus.

The ten topics, along with five of the most common words within each, presented alongside each topic label, were yielded using the STM modeling are presented below:

(1) Military

[veteran, military, air, guard, defense]

(2) Jobs and Businesses

[loan, business, companies, student, workforce]

(3) Education

[school, district, academic, student, attend]

(4) Budget

[spend, budget, debt, republican, deficit]

(5) Energy

[energy, oil, tax, gas, price]

(6) Government Accountability

[GAO (Government Accountability Office), audit, postal, inspector, review]

(7) Disadvantage Populations

[women, family, violence, children, american]

(8) Emergency Relief

[FEMA (Federal Emergency Management Agency), hurricane, disaster, repair, fund]

(9) Law and Judiciary

[amendment, constitution, bill, act, justice]

(10) Health

[health, insurance, care, patient, medicare]

Interpretation:

Figure 3b below presents an example of the clustering output from LDAvis. In this example, the algorithm selections used were *Jensen-Shannon* to calculate topic distance and *Sammon's Non-Linear Mapping* as Multidimensional Scaling Method. The number of clusters was 4, which seemed to be justifiable. For example, a plausible reason for the topics 'Education' and 'Disadvantaged Populations' being categorized within a single cluster may be due to the fact that these two topics showed signs of overlapping. Documents in these two topics both contained a significant number of similar words, such as '*children*' and '*arts*' leading them to be placed closely together.

Figure 3b: STM Visualization after Clustering of Topics

7.2 Topic Distribution by Frequency over Time

Motivation:

[1] Frequency of each word for each topic is defined as a relative frequency compared with the frequency of this word across all the documents.

This visualization seeks to interactively analyze the frequency trend of each topic along a time scale, thereby capturing the political interest in various topics presented in news releases over the time period of 2012. This also acts as a validation for the topic generation processes conducted beforehand, as we seek to determine the reasons for observed peaks and troughs that appear for each topic at various moments in the year.

Originality of Design:

We modified the original work of T. Craft (2012), in which he developed an interactive visualization of multiple area charts that utilizes a context tool to zoom and pan the electricity consumption per capita data from 1960 for France, Germany, Japan, UK and USA. It is suitable for our design purpose, which was to capture the trend of changes in frequency over time, for each of the topics obtained previously. The interactive component also enables us to analyze these trends in shorter durations of time. Here, the original yearly scale to alter to a monthly scale, which enabled us to present our frequency data in monthly intervals.

Aesthetics:

The order of these area charts is consistent with the order of topics we generated from STM model. We chose bright colors for each topic to make them more distinguishable to the end user.

Interpretation:

The visualization is presented in Figure 4 below. Upon rationalized the peaks and troughs for each topic category, we were able to observe the existence of interesting patterns that corresponded to the actual political events or national occurrences in 2012. For example, there were two peaks for the topic labeled 'Budget' – one occurred in February and the other was in December. The former one coincided with the annual budget release of $3.8 million, and the latter occurred in the time when heated debates regarding the fiscal cliff (whether the US federal government should extend the expiring tax cuts implemented by the Bush Administration) took place. The large spikes observed in the topic labeled as 'Emergency Relief' were associated with the occurrences of severe disasters, such as Southeast and Mid-West Tornados in February and North American derecho in June. The peak observed in June 2012 under the topic labeled 'Health' was attributed to the intense discussions surrounding the possibility of renewing the Obama Affordable Care Act.

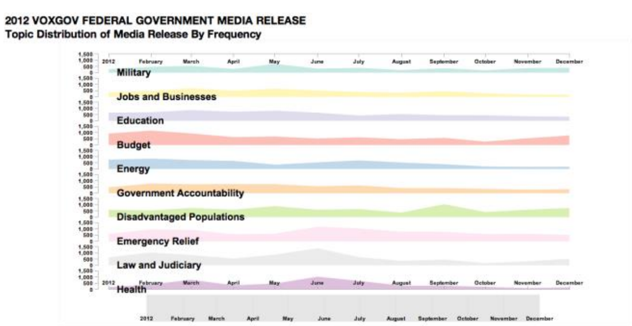

Figure 4: Topic Distribution by Topic over Time

7.3 Topical Static Sentiment Time Series Visualizer (Calendar Heatmap)

Motivation:

By using the calendar heatmap, we are able to present changes in daily sentiment scores for each topic across the year 2012. In this way, we were able to highlight the occurrences of certain events in the year, and discover the attached sentiment presented when these events were reported in the media.

Originality of Design:

In the original source prototype, the calendar heatmap is used to demonstrate daily financial historical data obtained from Yahoo finance. We incorporated the daily sentiment scores for each topic into the single year calendar for 2012, and adjusted the color scale ranging from dark red (extremely negative) to dark blue (extremely positive) to be consistent with all other visualizations.

Aesthetics:

In our visualization, the layout of calendar heatmap is the same as the original one. Moving from top to bottom and left to right, we present the sentiment scores corresponding to days of the week, from the months January to December respectively. Months are separated by thick dark outlines and days in each week are arranged into vertical columns. There are 11 color scales comprising five reds and five blues ranging in intensity for negative and positive sentiments respectively, along with a single white color to represent days with no media releases for that topic. A single calendar heatmap was created for each topic. This way, we display holistically the changes in sentiments that occurred in a daily basis across the ten generated topics.

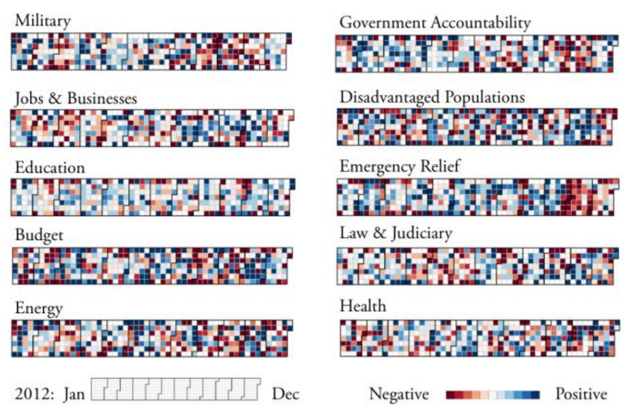

Figure 5a: Calendar Heatmap for all Topics

Interpretation:

Generally speaking, daily sentiment varies largely through the year for all topics, as seen in Figure 5a above. There is in general no clear trend that is revealed from the calendar heatmap, with the exception of the topic labelled 'Emergency Relief'.

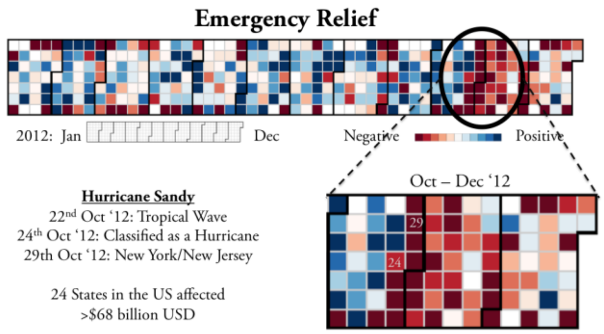

Figure 5b: Calendar Heatmap for 'Emergency Relief' Topic

As seen in Figure 5b above, there is a clear portion revealing negative sentiment in the ending months of October to November of the year 2012. Tracing back to see what happened during this period of time, we attributed the negative sentiment presented in media releases to Hurricane Sandy. Sandy developed from a tropical wave that formed on 22nd October 2012. By 24th October, Sandy was classified as a hurricane and was predicted to adversely impact several states in the United States. This was the main reason for the change in sentiment from positive to negative that occurred from 23rd to 24th October 2012. Hurricane Sandy affected 24 states and the overall damage was more than US$68 billion. This is consistent with the long-lasting, largely negative sentiment that followed in wake of the disaster. Following 24th October 2012, the darkest red (negative sentiment) lasted for about 10 days, reflecting the severity of the disaster.

7.4 Topical Interactive Sentiment Time Series Visualizer (Cubism)

Motivation:

This visualization, similar to the calendar heatmap, aims to present the trends of daily sentiment changes across each topic. The difference between the both, however, is that this interactive visualization modeled by the Cubism, a D3 plug-in, enables the precise comparison of sentiment scores across topics at one glance.

Originality of Design:

Cubism, a D3 plug-in first developed by Square, Inc. (2012), is primarily used for tracking real time data changes, such as financial stock pricing. We changed the rolling-time design to a static visualization, capturing only historical and not updated (current) data. This also involved the changing of steps and the visualization width to only show the period that we are interested in, as opposed to the original source.

Aesthetics:

To promote consistency across all our sentiment visualizations, we changed the default colors (green for positive numbers and blue for negative numbers) to blue for positive sentiment and red for negative sentiment. We included the ten names of the individual topic on the right top of each horizontal chart respectively. Mousing over each day, the corresponding sentiment score for each topic shows up on the left.

Interpretation:

For example, on the 29th October 2012 when Hurricane Sandy hit the northeastern U.S. coastline, the sentiment scores for the topics 'Emergency Relief' (-0.98), 'Health' (-1.3), 'Budget' (-1.6) and 'Job & Business' (-0.78) were consistently negative, as depicted in Figure 6. These results were reasonable in the direction of sign in sentiment scores, since a destructive disaster like such would require draw a large amount of media attention, focusing on expenditure to organize emergency efforts and healthcare, while also casting an overall negative outlook on the overall US economy. In line with the previous findings noted in the calendar heatmap, there is a persistence of negative sentiment being noted in 'Emergency Relief' in the ending months of 2012.

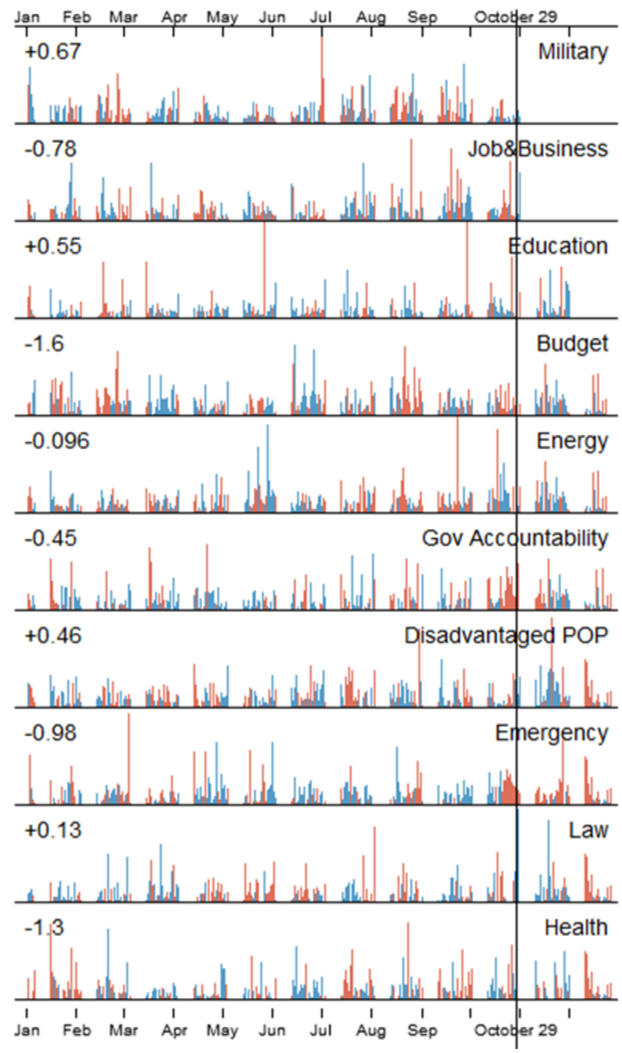

Figure 6: Cubism Visualization for Sentiment over Time

7.5 Topical Sentiment Benchmarking by Political Party and Gender

Motivation:

We would like to evaluate how sentiment scores varied according to the political affiliation as well as the gender of the respondent, for each topic in 2012. The following two static visualizations will enable us to present these findings suitably.

Originality of Design:

The two visualizations that benchmark topical sentiment according to political party affiliation and gender are original deliverables. Using ggplot from R to draw the basic graph, we separate the portions of the graph into positive and negative halves.

Aesthetics:

In both the visualizations, we divide the whole plot into two equal parts. The left portion would contain coordinates that express an overall negative sentiment, while the right portion would contain coordinates that express an overall positive sentiment for each topic. We colored the background of these halves with either blue or red, which indicates negative and positive and sentiments respectively. The y-axis represents the ten topics while the x-axis represents the overall sentiment score ranging from -2.0 to 2.0. The position of each icon corresponds to the simple dot plots produced from R. In our final version, we use icons to facilitate audiences to see which political party or gender the point represents.

Interpretation:

The sentiment scores we used for these visualizations were based on aggregated annual data. So, this showed a general picture of the difference between political party affiliation as well as gender for each topic.

Analyzing the differences in sentiment between political parties for topics in Figure 7a, we can see that both parties basically share the same signs in sentiment across all topics except for 'Military' and 'Emergency Relief.' On the overall, the Democrats also tend to be more positive across all the topics (except 'Law and Judiciary') compared with the Republicans. However, the disparity between the two parties is not very apparent across topics, reflecting the fact that most of the scores from both parties lie fairly close to one another.

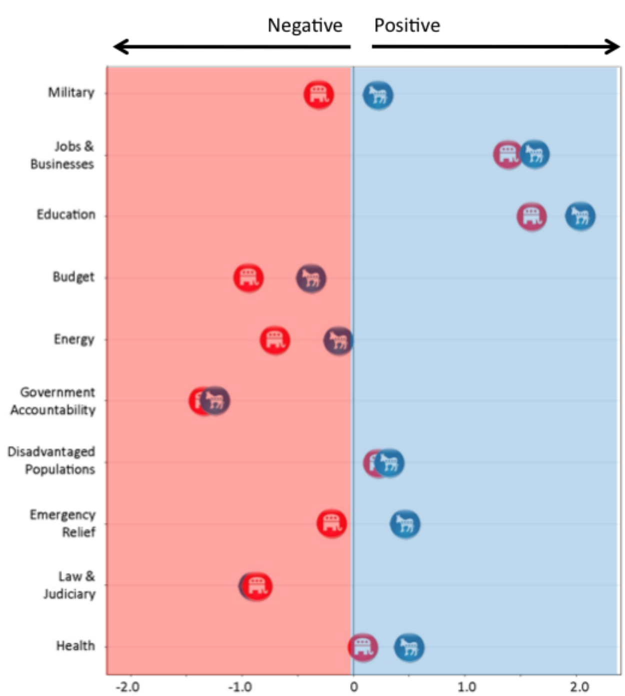

Figure 7a: Topical Sentiment Benchmarker by Political Party

Sentiment benchmarking based on the genders of respondents was conducted in a similar manner. As compared to political party affiliation, there is no clear trend if either males or females tend to be more positive or negative in terms of sentiment, as seen in Figure 7b.

However, upon taking a closer look between both visualizations, we note that the sentiments of respondents, either based on political party affiliation or gender, tend to line up consistently around the same value for each topic. These two visualizations thus mutually validate one another in terms of sentiment scoring across topics. It also kind of reveals the fact that official government release attempted to be as neutral as possible across different stakeholders.

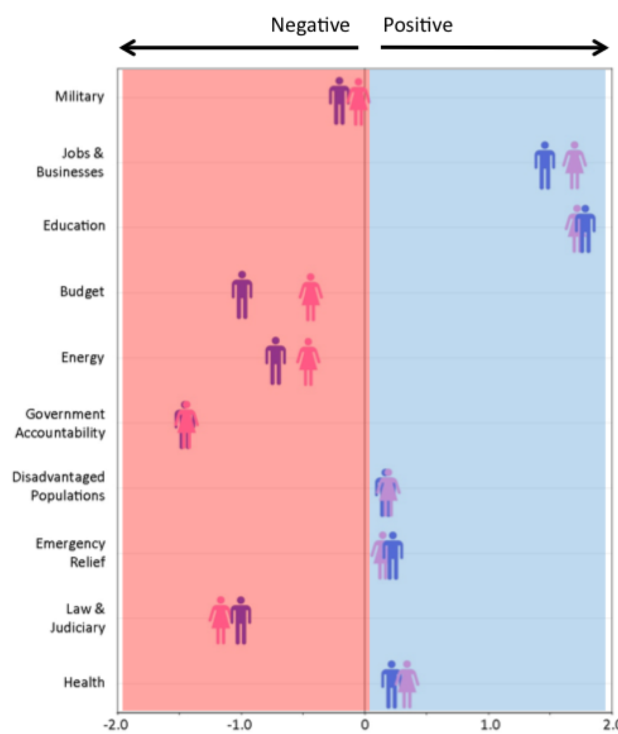

Figure 7b: Topical Sentiment Benchmarker by Gender of Respondent

7.5 Topical Sentiment Benchmarking by Location (Cartography)

Motivation:

We would like to see how sentiment varies across locations in the US, for different topics. In order to analyze location disparities, we used cartographical visualizations to highlight these differences over geographical space. This is conducted at the state level.

Originality of Design:

To facilitate the comparison across different topics, we used the same scale for the sentiment scores for each map. We also included the territories which were put in the boxes under the main US map.

Aesthetics:

Every single US map was generated to visualize the mean annual sentiment for each topic at the state and territory level. The color palette is the same as the one used in the calendar heatmap visualization, with the only exception that no white color existed since all states had corresponding sentiment scores for each topic. The blue colors represent positive sentiment while the red colors represent negative. The darker the color is, the higher the sentiment score. The states and territories were labeled so end users can easily identify each geographic location.

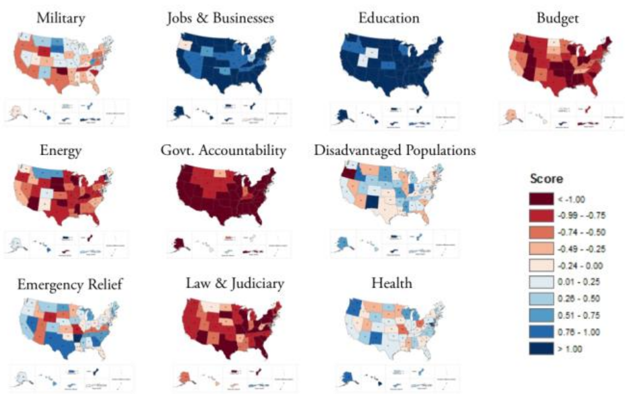

Figure 8a: Topical Sentiment Benchmarking by Geographic Location

Interpretation:

Again, we used the aggregated annual data in this visualization presented in Figure 8a. Across the ten separate maps, we can see that there are five topics whereby all locations share one overall political sentiment, either positive or negative, in media releases. These include the topics labeled 'Jobs and Businesses', 'Education', 'Budget', 'Government Accountability' and 'Law and Judiciary.' The other maps reveal that other topics are more varied in sentiment, highlighting that respondents from different locations may contribute different slants towards a given topic.

Due to the fact that annual data was aggregated and presented in these maps, it is difficult to trace the reasons why some locations show more intensity in sentiment scores than others. Additionally, the sentiment scores in one state within a year might cancel off due to multiple positive and negative affairs. Furthermore, certain news releases originating from a given state or territory may also be discussing occurrences that are happening outside of its state boundaries.

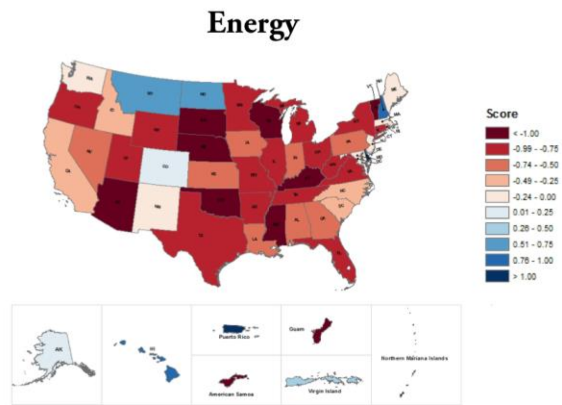

Figure 8b: Sentiment Benchmarking by Geographic Location for Topic 'Energy'

However, taking a closer look at the energy map as displayed in Figure 8b, we may be able to reason why the states of Montana and North Dakota report, on average, more positive sentiment in news releases as compared to other states in the US. These two states are primary oil and gas producers in the country. Having seemingly more abundant resources and more intensively involved in energy production as compared to the other US states, it is no wonder why media releases around the topic of 'Energy' would be so positive for both two.

Such analysis may also be conducted for other topics. Tracing back to the text files within the corpus that were clustered under each topical category, in accordance to their location sources, will also reveal insights for why certain states revealed sentiment patterns similarly or differently.

8. DISCUSSION AND FUTURE STEPS

The limitations and areas for future development are listed accordingly in three main topics. Firstly, we address the issues surrounding the STM process and the implications arising from that. Secondly, we highlight possible areas to improve sentiment analysis processes, mainly focusing on the procedure of scoring text documents in order to determine absolute and measureable sentiment scores. Lastly, we discuss the possible areas of improvement in some of the visualizations presented earlier.

8.1 Structural Topic Modeling

There are two main issues that surround our structural topic model that requires addressing.

First, the number of topics we chose, which was ten, lacked theoretical grounds. This was an arbitral number that seemed robust and distinct enough for the given research project. However, if given more time and resources, it is possible for us to try a larger number of topics to see which level was most suitable future analysis. Alternatively, this issue may be addressed by adopting more technical approaches, such as cross-validation, nonparametric mixture priors or using marginal likelihood put forth by M.A. Taddy (2011), so as to more aptly determine the optimal number of our topics.

Second, there exist certain issues surrounding the specification of parameters used in the STM procedure. For example, we used topical prevalence to include our metadata into our model because we hypothesized that it might be more reasonable to allow the metadata (party and gender) to affect how frequent a topic was discussed. However, it was also highly possible that these metadata affected how the words were used within one particular topic. This suggests that there exists another plausible option to specify our model via topical content. In the future research, the specification of such parameters might require more solid justifications.

8.2 Sentiment Analysis

A common way to calculate sentiment score of an article is to count the number of negative and positive words contained within it. Here, positive words are scored one point while negative words are scored minus one point each. The final sentiment score is the sum score of all positive and negative scores.

In our study, we implemented this method and further standardized the scores for each document by dividing the number of total positive and negative words to better compare among documents. However, there remains a potential limitation in using such method. It is possible to misjudge or obtain wrong sentiment scores for sentences or text paragraphs, since the sentiment lexicon is adjective-sensitive and neglects sentence structures (such as double negatives).

The recursive deep model by professors from Stanford University deals with this problem by utilizing an advanced algorithm to predict sentiments of articles. This model builds up a representation of whole sentences based on the sentence structure. It computes the sentiment based on how words compose the meaning of longer phrases, supported by the use of a Sentiment Treebank and Recursive Neural Tensor Network to validate the model. To improve our sentiment prediction system, we could apply this method to our further study and reduce any biases caused by neglecting sentence structure.

8.3 Visualization Improvements

As for now, our visualizations are separate and only show one aspect of different topics. It is, as a result, possible for us to incorporate more interactive components to the existing static visualizations in order to make the final deliverables more end-user friendly.

For example, in the cartography visualizations, we can only see the singular sentiment of regions for different topics. As such, it is not possible for us to present more detailed information (such as sentiments over time and location) regarding a given topic in a given state using our current visualizations. Inspired by trulia.com, an improvement we can conduct is to combine aspects of the calendar heatmap and cartography visualizations together, so as to provide more detailed information with the use of a single and interactive visualization. In this ideal visualization, we can click on each state on the US map and the related calendar heatmap will change accordingly. Furthermore, incorporating a bar chart next to the cartography can also reveal the daily frequency of each topic mentioned in each state. This way, end-users can not only see a general picture of sentiment across states but also see how daily sentiment changes in each state for each topic.

Another example is the limitations surrounding the Cubism visualization. Cubism is primarily designed for real-time tracking, and as a result there still exist some difficulty in modifying the default one-pixel per unit-time setting, so as to make the prototyped visualization more customizable. Moving forward, it is also ideal to change the horizontal chart from "mirror" (i.e. having both the negative and positive scores portrayed on the same axis) to "offset" (i.e. having the negative and positive scores on two separate, diverging axes) setting, so as to better highlight the differences between the negative and positive sentiment scores within each topic.

9. CONCLUSION

Compared to textual data sourced from social media platforms such as Twitter and Facebook, government releases are more formal and embedded with less social linguistics operators (such as hash tags and emoticons), making sentiment analysis more reliable and stable.

In this paper, we first proposed a structural topic model to categorize our political corpus into to ten major topics. In particular, we allowed the metadata, namely, party and gender, to affect the frequency one topic was discussed. The sentiment scores of documents (and topics) were extracted using a combined dictionary from various lexicon sources, and standardized thereafter in order to conduct various analyses.

In order to better deliver our results, we utilized D3, Shiny and ArcMap applications, amongst other tools, to create and deliver several static and interactive visualizations.

We find that political sentiment along time trends and geographical distribution for each topic lay fairly consistent with the occurrences of the events within 2012. Also, as initially postulated, the sentiments embodied in government media releases were relatively neutral and convergent across the paradigms of gender and political party affiliation.

Lastly, we note that our visualizations provide potential avenues for Voxgov to analyze, manage and present their massive database of media and news documents. For example, the visualization utilizing the Cubism package, having a real-time management capability, may be utilized by Voxgov to visualize political sentiments of the incoming media documentations that it collects on a continuous daily basis.

10. ACKNOWLEDGEMENT

We would like to express our gratitude to Professor Sharon Hsiao and Professor Greg Eirich for their patient guidance, enthusiastic encouragement and useful critiques of this research work. We would also like to thank Voxgov for kindly providing the dataset we used for our research purpose. We also gratefully extend our appreciation to Mr. Carson Sievert and Mr. Kenneth Shirley, LDAvis developers, who helped us better utilize the LDAvis package, and also Mr. Kai S. Chang, D3 Parallel Coordinate developer, for his technical support in our Cubism visualization.

11. REFERENCES

[1] Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier.*NAACL 2013*, 49.

[2] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.

[3] Bora, N. N. (2012). Summarizing public opinions in tweets. *Journal Proceedings of CICLing*.

[4] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

[5] CNN Library (July 13, 2013). Hurricane Sandy Fast Facts. *CNN World*. Retrieved from http://www.cnn.com/2013/07/13/world/americas/hurricane-sandy-fast-facts/

[6] Craft, T. (2012, August 29) Multiple Area Charts with D3.js. [Web log post] Retrieved from http://tympanus.net/codrops/2012/08/29/multiple-area-charts-with-d3-js

[7] Davidov, D., Tsur, O., & Rappoport, A. (2010, August). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 241-249). Association for Computational Linguistics.

[8] Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[9] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.

[10] House Hunting All Day, Every Day. (n.d.). In *Trulia*. Retrieved May 10, 2014, from http://www.trulia.com/vis/tru247/

[11] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.

[12] Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2, 627-666.

[13] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series.*ICWSM*, *11*, 122-129.

[14] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*(pp. 79-86). Association for Computational Linguistics.

[15] Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC*.

[16] Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. *Austin, TX, LIWC. Net*.

[17] Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2013). Structural Topic Models. Working paper.

[18] Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3-55.

[19] Sievert, C. and Shirley, K. (June, 2014). LDAvis: A method for visualizing and interpreting topics. Paper presented at *ACL2014 Workshop on Interactive Language Learning, Visualization*, and Interfaces, Baltimore, MD.

[20] Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1631-1642).

[21] Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440.

[22] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. *In Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.

[23] Taddy, M. A. (2011). On estimation and selection for topic models. *arXiv preprint arXiv*:1109.4518.

[24] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.*ICWSM*, *10*, 178-185.

[25] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *In Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.

.