

Final Paper: Topic Modelling of Inaugural Speeches

Yanik L. Kipfer

04 Januar, 2020

1. Introduction

George Washington held the first inaugural speech during his inauguration as the first US president in 1789. Ever since then, most presidents have upheld the tradition of addressing the nation, during their presidential inauguration, through a speech. Within these speeches, presidents tend to outline the challenges that the nation faces and how they plan on solving them during their term of office. Additionally, the speech is also meant to unite and ensure the support of the population for the new administration, after the election is over. Thus, inaugural speeches can contain valuable information on the most salient topics during the election, as well as the president's agenda.

A closer analysis of the inaugural addresses delivered by the various presidents might, therefore, deliver valuable insights on which presidents addressed similar topics in their speeches and hint at the salient issues they faced during their terms of office.

The primary aim of this paper lies in the attempt of identifying the main topics contained within the various inaugural speeches and subsequently, identify the similarities and differences between the speeches.

Following, I will first describe the data used for the analysis, the method employed to identify the various topics, to then present the results and conclude on the findings.

2. Data

The original data is comprised of 59 presidential inauguration speeches from 36 presidents ranging from the year 1789 to 2017. A full list of all presidents with inauguration speeches present in the dataset can be found in at the end of the paper.

In order to perform the analysis, the dataset is converted into a document-term matrix. After removing all the punctuation, spacing between words, as well as commonly used words (so-called stopwords) and very rare words (here I have set the sparse factor to 0.90, removing all words that occur less than 10% in all documents) which do not contribute any information for the topic modelling process, the final dataset spans over 59 document and 1949 terms.

Looking at the total term distribution within the document-term matrix (Figure 1), makes it apparent that inaugural speeches tend to extensively use the terms *country*, *world* and *nation*, since these terms can be found over 300 times in the data.

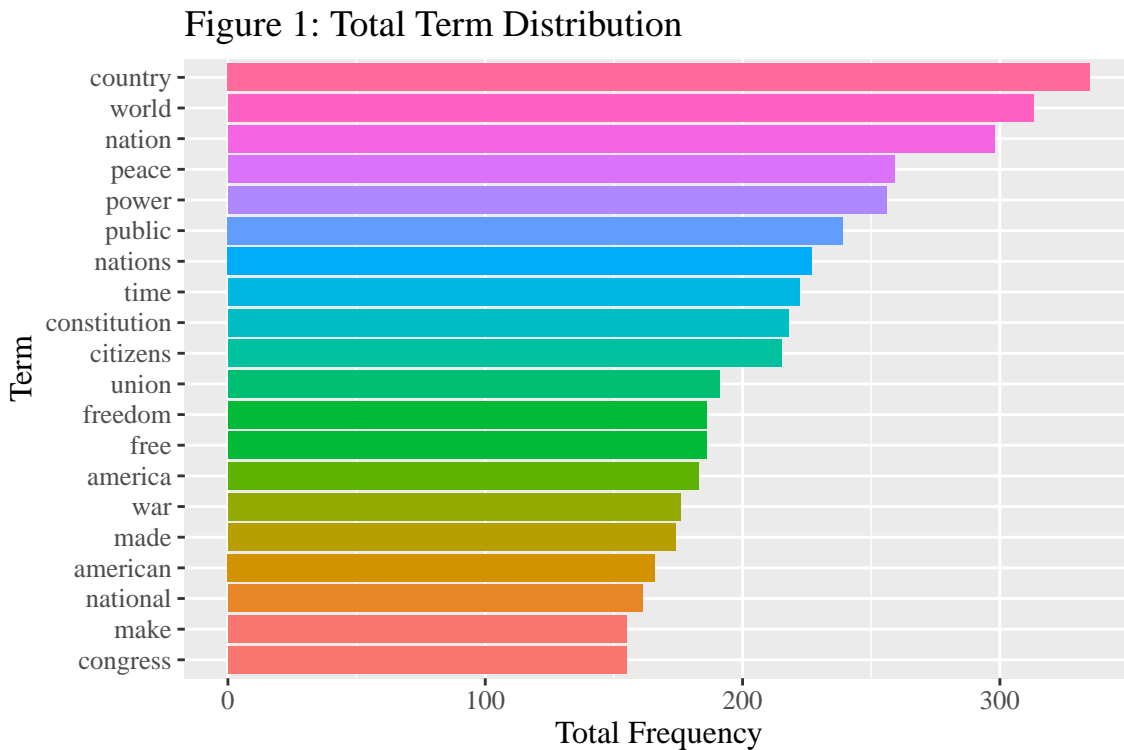


Table 1: Explained Variance of Topics

	Explained Variance	CumSum
Topic 1	41.6	41.6
Topic 2	9.2	50.8
Topic 3	4.7	55.5
Topic 4	3.1	58.6
Topic 5	2.9	61.5
Topic 6	2.3	63.8
Topic 7	2.0	65.8
Topic 8	1.9	67.7
Topic 9	1.7	69.4

3. Method

In order to extract the primary topics contained within the speeches, I use **Non-Negative Matrix Factorization** (NMF). Similar to Singular Value Decomposition (SVD), NMF is a dimensionality reduction method with the main difference between SVD and NMF lying in the fact that NMF accounts for the non-negative nature of text data, by constraining the values within the matrices to only take on non-negative values.

NMF seeks to extract the latent topics in the document-term matrix by decomposing it into a topic-term and document-topic matrix. The topic-term matrix delivers information on the latent topics found within our data, while the document-topic matrix tells us about the association between the documents and the latent topics. The main idea behind the decomposition is that when the topic-term and document-topic matrices are multiplied, they should approximately reconstruct our original document-term matrix. To achieve the best approximation of our original document-term matrix, NMF tries to minimize the loss of information caused by the decomposition, through an iterative process.

To perform NMF, we first need to define the number of topics the algorithm should extract. Using a eigen decomposition and a explained variance threshold of 70%, table 1 shows that 9 topics should be extracted to be able to explain 70% of the total variance.

4. Results

In the following section, I will present the results of the topic modelling analysis. For the sake of easier comprehension, I will first show and briefly interpret the latent topics found by NMF and, subsequently, look at the association between these topics and the presidential speeches.

4.1. Term-Topic Distribution

Figure 2 shows the distribution of the top ten terms for each topic. Starting with the **first topic** we see that amongst its most frequent terms used are words, such as world, America/American, nation and freedom. Thus, overall topic 1 seems to mostly address **America as a nation**.

The **second topic** is heavily characterised by words like public, union, foreign, nations and interests. Therefore, it can be interpreted to pick up on **issues of foreign policy**.

Turning to the **third topic** it immediately comes to attention that this topic speaks to themes concerned with war, congress and country, thereby, it might point to issues of **civil conflict/war**.

Topic 4 can be interpreted to pick up on the theme of **constitution and the union**, while **topic 5** deals with issues of **world peace**.

Topic 6 similar to topic 4 seems to pick up on constitutional issues; however, it seems with more of a focus and institutions and the executive. Therefore, it can be interpreted on picking up on the topic of **separation of powers**.

Topic 7 concerns itself with the topic of **law**, while **topic 8** is harder to interpret, looking at the term distribution we see that the terms most used in topic 8 are nation, life, men, things, good and spirit. These words point to uplifting themes and, thus, might address the topic of *nation-building*. Finally, **topic 9** picks up on topics concerned with the **economy**.

Table 2 provides a full overview of all the identified themes for all topics.

Figure 2: Topic–Term Distribution By Topic

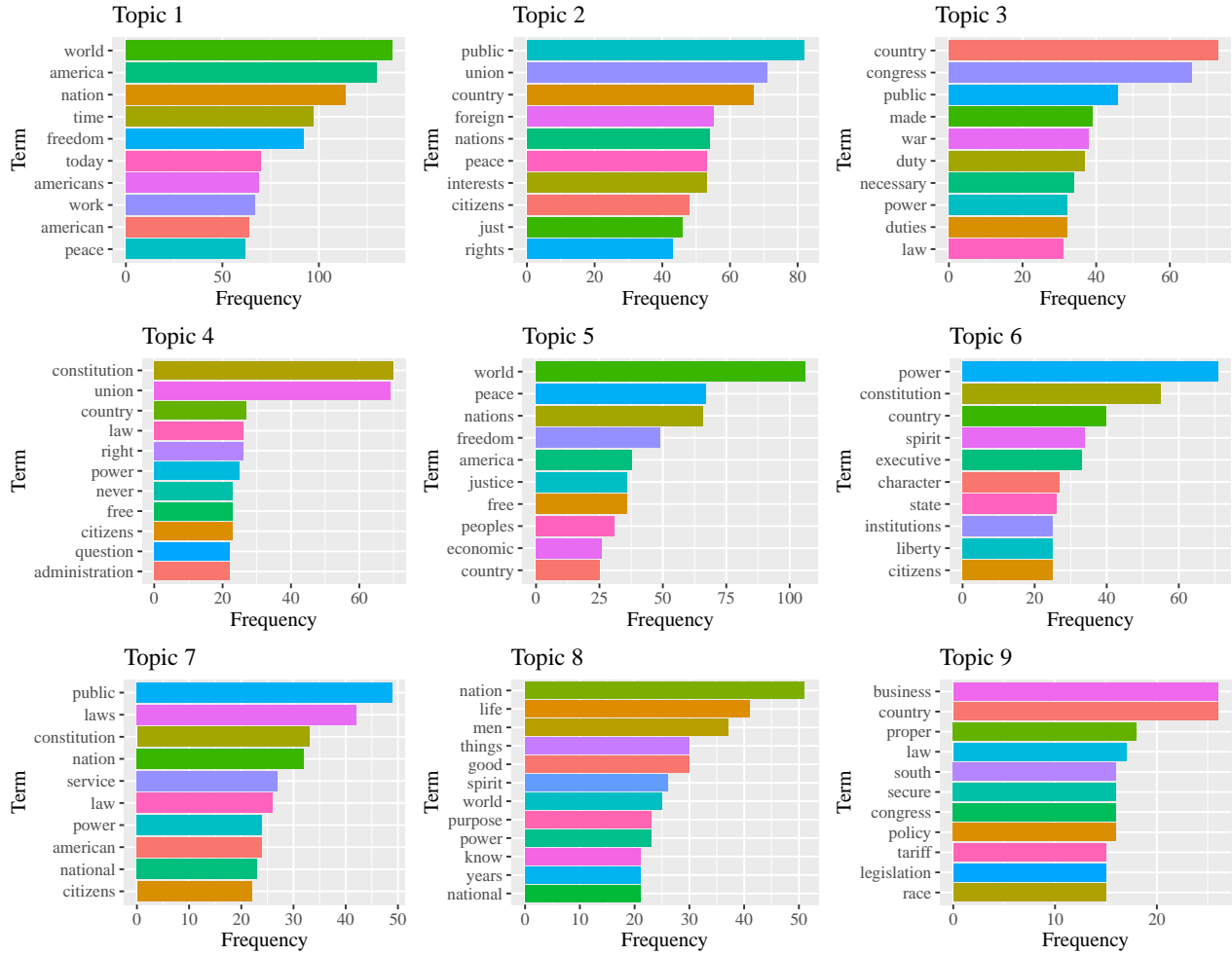


Table 2: Topics And Their Themes

topics	themes
Topic 1	America As A Nation
Topic 2	Foreign Policy
Topic 3	Civil Conflict
Topic 4	Constitution & the Union
Topic 5	World Peace
Topic 6	Seperation of Powers
Topic 7	Law
Topic 8	Nation Building
Topic 9	Economy

4.2. Document-Topic Distribution

This section will, finally, look at how strongly the different speeches are associated with the previously identified topics and elaborate on the most interesting patterns found through these associations.

Figure 3 presents the document-topic distribution and shows us the 10 documents that are most strongly associated with a particular topic. One thing that becomes immediately apparent when looking at the speech-topic distribution is that speeches from presidents that had their terms in office relatively close to each other tended to address the same topics. Looking at Topic 1, we see that the speeches that are most strongly associated with this topic are all from presidents who served within the last 50 years. During this time, Americans experienced the fall of communism, the terrorist attacks on September 11th 2001 and, with Barack Obama, the first black president in American history. The speeches made during these times, thus, appealed greatly to the notion of being an American and an American nation. Similar can be said for topic 8, which is most strongly associated with speeches at the beginning of the 20th century. The beginning of the 20th century were times of great tumult. America and the world experienced two world wars and the Great Depression. Thus, it is not suprising that these speeches tended to have more uplifting sentiments, as contained in topic 8.

Therefore, while it is not surprising that similar issues confronted presidents who served close to each other, the different topics found through NMF show that the salient topics changed over time.

Furthermore, we can see that the previous interpretation of the topics is mostly consistent with the speeches associated with them. Focusing on topic 4, we had previously stated that this topic addressed themes to do with the constitution and the Union. President Lincoln's speech is most strongly associated with topic 4. Lincoln served as a president during the American Civil War, a time where issues of the constitution and the Union of states were most pressing in American politics.

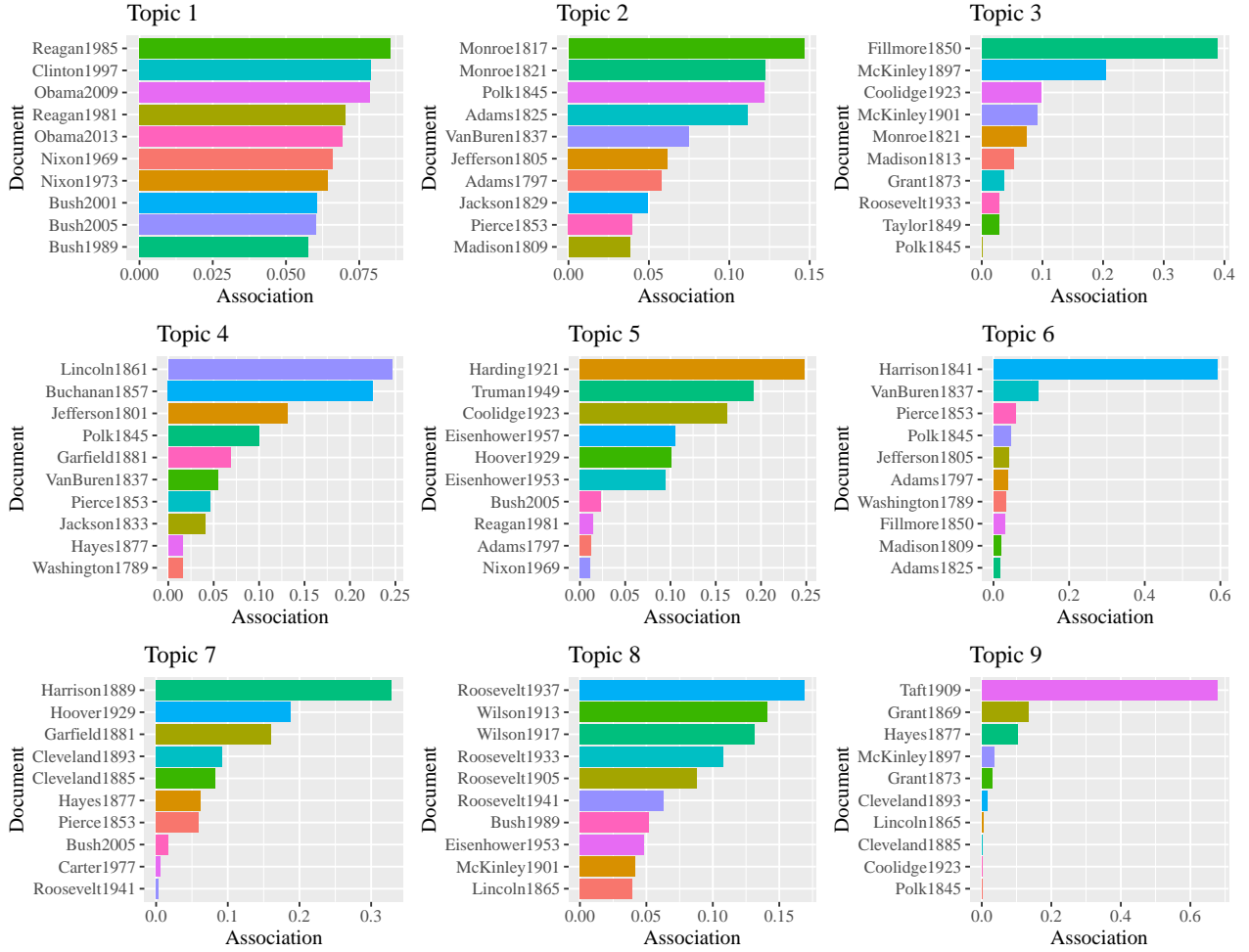
Furthermore, Topic 2, which mostly picks up on topics of foreign policy, also seems to fit the speeches it is strongly associated with. On the one hand, European colonialism of the Americas was a central issue during Monroe's term in office and his inaugural speech. Monroe was a staunch opponent of European colonialism in the Americas and, through the Monroe Doctrine, Monroe determined that the Americas were to be independent of European influence, which was a key moment in US foreign policy. On the other hand, during his inaugural speech, Polk talks at length on the foreign policies issues. In specific, the main foreign policy issue addressed by Polk was his desire that Texas would rejoin the Union. Up till that point, Texas was still formally a part of Mexico.

Another example of the congruency between topic and speeches can be seen when looking at topic 5. The speeches most strongly associated with topic 5 are mostly from presidents who served shortly after the first or second world war. Unsurprisingly, therefore, topic 5 mostly speaks to themes of world peace. When taking a closer look at the speeches made by Harding, Coolidge and Truman, all three heavily address the topic of peace.

Overall, we can say that NMF seems to have correctly identified the different topics addressed by the inaugural speeches and that the most striking pattern found within our data is that presidents who served close to each other, addressed similar topic within their speeches and

that these topics changed over time.

Figure 3: Topic–Document Distribution by Topic



5. Conclusion

This paper set out to identify the different topics contained within presidential inauguration speeches and, thereby, find patterns amongst these speeches. Using data on 59 inaugural speeches from 36 US presidents, 9 topics were identified and extracted through Non-Negative Matrix Factorization. Through the interpretation of the term distribution amongst the 9 topics, I identified that these topics addressed different issues reaching from themes concerned

with world peace to such concerned with civil conflict.

Subsequently, I took a look at which speeches were most closely associated with the different topics and tried to find patterns within this distribution. I concluded that presidents who served their terms in the office close to each other tended to address the same topics.

I also took a look at how well the topics identified the inherent themes contained within the speeches, by comparing the content of a few speeches to the interpretation of the topics they were most associated with. Taken as a whole, the interpretation of the topics were a close fit to the content of the speeches.

Table 3: List of Presidents With Inauguration Speeches

year	president	year	president
1789	Washington	1905	Roosevelt
1793	Washington	1909	Taft
1797	Adams	1913	Wilson
1801	Jefferson	1917	Wilson
1805	Jefferson	1921	Harding
1809	Madison	1923	Coolidge
1813	Madison	1929	Hoover
1817	Monroe	1933	Roosevelt
1821	Monroe	1937	Roosevelt
1825	Adams	1941	Roosevelt
1829	Jackson	1945	Roosevelt
1833	Jackson	1949	Truman
1837	VanBuren	1953	Eisenhower
1841	Harrison	1957	Eisenhower
1845	Polk	1961	Kennedy
1849	Taylor	1965	Johnson
1850	Fillmore	1969	Nixon
1853	Pierce	1973	Nixon
1857	Buchanan	1977	Carter
1861	Lincoln	1981	Reagan
1865	Lincoln	1985	Reagan
1869	Grant	1989	Bush
1873	Grant	1993	Clinton
1877	Hayes	1997	Clinton
1881	Garfield	2001	Bush
1885	Cleveland	2005	Bush
1889	Harrison	2009	Obama
1893	Cleveland	2013	Obama
1897	McKinley	2017	Trump
1901	McKinley		