Assessing U.S. Politics Through 234 Years of Presidential Speeches

Eric Andrews School of Information
Nate Tran School of Information
Kaleb Parker School of Information
Matthew Schneider Rackham Graduate School

Project Motivation

- United States history has seen substantial changes in not only policy, but also attitudes towards varying countries/ethnic groups
- The national lexicon to communicate these attitudes has also changed
- We sought to quantify these changes in attitude through Natural Language Processing (NLP) techniques, specifically Word Embeddings & Sentiment Analysis

Previous Literature

- Word embedding work has captured increasingly positive connotations surrounding immigration despite increasing political polarization (Card et al., 2022)
- Sentiment Analysis has been an effective method for capturing collective shifts in real time during political debates (Wang et al., 2012)

Questions

1. How have word similarities changed over time when considering ethnic/gender based terms?
2. Is there any relationship between the terminology used and the general sentiment during that time period in the United States?
3. How does sentiment/word embeddings vary when considering different subsets of the data (i.e. Political Party, Era in American History, etc.)?

Methods

Preprocessing:

- Clean speech transcript data to standardize capitalization, remove punctuation and stopwords

Word Embeddings:

- Split data into 20 year eras (about 5 presidential terms) to create models over time
- Use Doc2Vec to train word embedding models based on each 20 year split
- Compare most similar words using cosine similarities based on keyword over each 20 year model
- Identify important trends amongst similar words over time

Sentiment Analysis:

- Utilize Spacy package to create NLP model
- Use Sentiment Intensity Analyzer from the NLTK sentiment vader package to generate sentiment scores
- Split data by party and 40 year eras
- Calculate running average sentiment within each party
- Calculate average sentiment for all parties in each era
- Create visualizations to plot running average sentiment over time

Results

Word Embeddings

Enemy
Similar Words Over Time

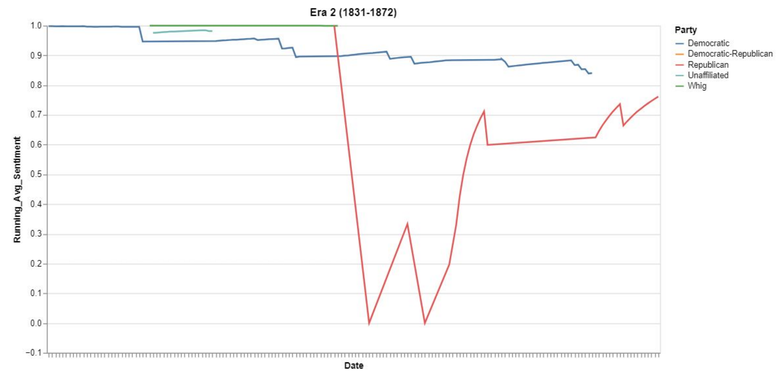
Word	Time Period
"Canada" / "Canadians"	War of 1812
"Mindanao"	Battle Fought in Mindanao, Philippines in WW2
"Naval" / "Fleets"	WW2
"Saddam" / "Hussein" / "Dictator"	During Hussein's Iraqi Reign
"Laden"	9/11

Slavery
Similar Words Over Time

Word	Time Period
"instrumental"	Before 13th Amendment
"accepts"	Before 13th Amendment
"participating"	Before 13th Amendment
"miseries"	Past 20 Years
"extreme"	Past 20 Years
"cruelty"	Past 20 Years

Sentiment Analysis

Party	*Average Sentiment*
Era 1 (1789-1830)	
Unaffiliated	0.9744
Federalist	0.9934
Democratic-Republican	0.9587
Democratic	0.9981
Era 2 (1831-1872)	
Unaffiliated	0.9907
Whig	0.9996
Democratic-Republican	0.6783
Democratic	0.8308
Republican	0.7614
Era 3 (1873-1914)	
Democratic	0.9035
Republican	0.8898
Era 4 (1915-1956)	
Democratic	0.7886
Republican	0.9223
Era 5 (1957-2019)	
Democratic	0.9187
Republican	0.8543



Discussion / Conclusions

- Word Embeddings follow trend of ongoing U.S. conflict/events throughout history.
- The era in American History plays an important role in word embedding trends depending on current state of the country.
- No clear correlation between party affiliation and sentiment
- Results of sentiment analysis are most interesting when observing differences in sentiment across eras throughout history
- Timing of large dips in average sentiment line up with major events such as the Civil War and the Great Depression.