Tracking Political Sentiment on Cold War China in Congressional Speeches

Faculty Advisor: Zach Branson

By: Zachary Novack, Eden Hu, Mason Lin

External Advisor: Dani Nedal

Introduction

- The China–US relationship has been described as one of the most important bilateral relationships of the 21st century. Richard Nixon's Presidency helped to strengthen diplomatic ties with China at the heart of the Cold War and thus reshaped US and global foreign policy.

- In this project, we analyzed nearly 2 million Congressional speeches from 1946 to 1991 to examine how political sentiment on China has evolved over time. Using topic modeling, sentiment analysis, and other text analytic methods, we also assessed how external political events have triggered large-scale shifts in opinion on US-China relations.

Data / Methods

Data Processing:

- Our data consists of 1,890,903 congressional speeches from 1946 to 1991
- To pinpoint Congressional speeches that do and do not discuss China, we split the data into three pieces:

1. **Speeches that don't mention China -** Speeches that contain no words from dictionary of pre-identified China words
2. **Subspeeches that mention China -** The parts of speeches that mention China according to the following heuristic:
   - For each "China word" within a speech, select that sentence AND the two sentences before and after
3. **Subspeeches that don't mention China in a speech that does mention China -** The parts of a speech that mentions China that are not part of (2)

How am I doing today? I'm doing quite well, thank you for asking. I just recently returned from my trip abroad. China was such an amazing place. You should definitely go if you can. I can't wait to go back in a few years. I'm pretty jetlagged from the flight though. The airport was also a whole mess. There were crying kids everywhere. Also, all that Chinese food I had didn't sit right with me on the flight. Hopefully I feel better soon.

Example of windowing approach; sentences that mention China (red), sentences around each China mention (yellow), and non-China sections (white)

Feature	Description
`text` (string)	Text of actual speech or subspeech (stopwords removed)
`speech_id` (int)	Unique numeric ID for each speech
`party` (string)	Political party (e.g. Democrat, Republican)
`chamber` (string)	Body of Congress (House or Senate)
`date` (Date)	Date of speech
`ave_sentiment` (float)	Average +/- sentiment score for given speech (or subspeech)
`emotion` (float array)	Vector of probabilities that a speech (or subspeech) conveys a specific emotion
`topic` (float array)	Vector of probabilities that a speech (or subspeech) belongs to a given topic

Table of existing (green) and estimated (orange) features from the dataset

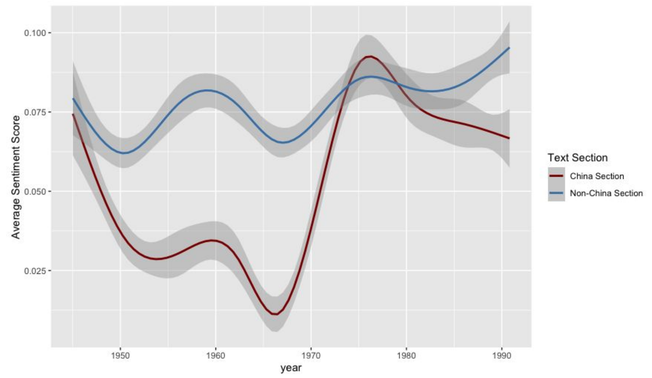
Methods:

We used three different techniques to estimate content of each speech or subspeech:

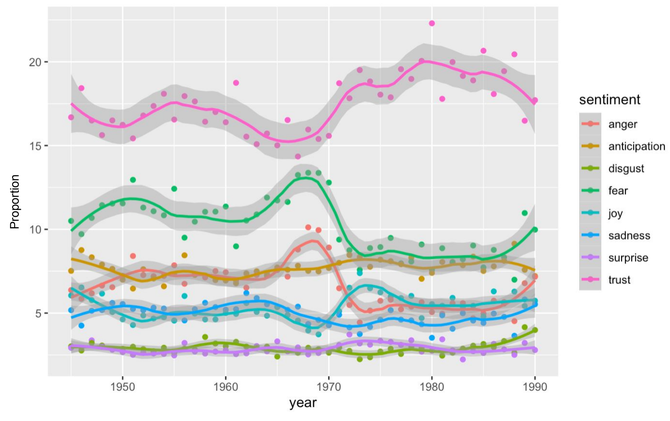
1. **Valence-Adjusted Sentiment[1]:** Estimates positive/negative sentiment of a text but accounts for valence-shifters (e.g. "not", "isn't", etc.)
2. **Emotion Categorization[2]:** Estimates probability that text conveys a specific emotion (e.g. joy, fear, trust)
3. **Latent Dirichlet Allocation[3] (LDA):** Generates a topic model of the data to estimate the probability that a text talks about a given subject

Analysis & Results

Binary Sentiment Analysis



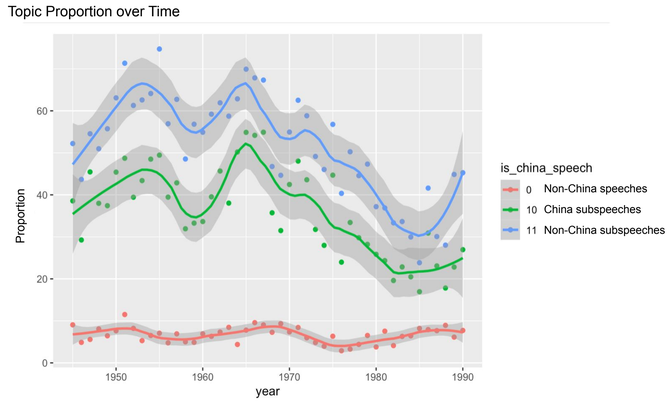
Emotion Categorization



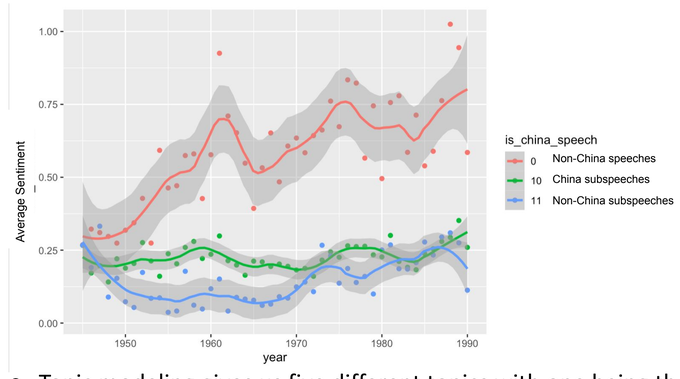
- Higher value in the top plot means more positive sentiment. There is a positive sentiment change around the time Nixon visited China. This positive sentiment change could be explained by the increase in trust and joy words as well as the decrease in fear and anger words.

Topic Modeling (with k = 5) & Sentiment Analysis

Vietnam and Korean Wars



Sentiment over Time



- Topic modeling gives us five different topics with one being the Korean and Vietnam wars.
- This topic in particular demonstrates that the US-China relationship was also impacted by many of the external factors.
- The China portions in the China speeches talked about the two wars the most often and have the lowest sentiment during that period.
- Not shown in the graph, but the key emotions that are in sync with topic proportion are anger and fear.

Conclusion

- Using both +/- and emotion-based sentiment analysis, we found that there exists a significant "growth" period during Nixon's presidency where sentiment on China turns more positive.
- Subspeeches about China are less positive than the surrounding text throughout this period.
- Topic modeling on the speeches failed to identify any clear topics during this "growth" period; however, it did identify topics like the Korean and Vietnam wars, and we found that the sentiment toward war topics is lower for subspeeches that are about China than other speeches.
- Controlling for cofactors (Political Party, Congressional Chamber) did not yield any significant differences

References

1. Rinker, T. (2019, March 22). Package 'sentimentr'. sentimentr documentation. https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf.
2. Saif, M. (2013). NRC Emotion Lexicon. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.
3. David, B., Andrew, N., & Michael, J. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993–1022.