Assignment 1: What's the Mood? Unveiling the Sentiments in the Debate

Research Report

Mai Ngoc Nong

Department of Network and Data Science

DNDS6014: Introduction to Computational Social Science

Prof. Mark Wittek

October 18, 2024

1. Introduction

The 2024 U.S. Presidential Election represents a critical moment on a national and worldwide level alike. With Vice President Kamala Harris representing the Democratic Party and Former President Donald Trump representing the Republicans, the race has ignited intense debates over issues such as the economy, abortion, immigration, and foreign policies. This research report aims to uncover sentiments expressed by both Harris and Trump during the debate held on September 10, 2024, at the National Constitution Center in Philadelphia through emotion and sentiment analysis on the debate transcript. Through an overall analysis of polarity, subjectivity, and emotions, the key differences between Harris' and Trump's sentiment at a high level are revealed. Additionally, this research also examines sentiment and emotions expressed at the level of specific issues raised during the debate, highlighting how sentiments evolve over time and how they resonate emotionally with voters, all while reflecting their party ideologies.

2. Data Preprocessing

The overall data preprocessing pipeline starts by separating the transcripts for each candidate, followed by a sequence of text-cleaning steps, including converting to lowercase letters, removing punctuation, tokenizing, and removing stop words.

For temporal analysis, since the goal is to measure sentiment and emotions expressed through the speeches chronologically, each candidate's speeches are stored in a list rather than merged into one text string. Further text-cleaning procedures are applied to each speech individually.

As for deep-dive analysis of each topic, the topics and relevant transcript sections are extracted manually by examining the transcript, paying attention to the hosts' announcement every time a new topic is introduced. The text cleaning process for sentiment analysis by topic follows the same pipeline as the overall analysis.

3. Data Analysis

Taking a top-down approach, the analysis begins by exploring the overall sentiment and emotion. Using the preprocessed transcript of each candidate as input, polarity and subjectivity scores were measured with the TextBlob library. Polarity, measured on a range of [-1, 1], represents negative or positive sentiment. Subjectivity ranges between 0 and 1 with zero meaning objective and 1 meaning subjective opinion.

Next, the sentiment strength, or intensity, is assessed with nltk's sentiment analysis module, which is a part of the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analyzer. The output provides proportions of positive, negative, and neutral sentiment in the text that add up to 1, along with a compound score, which is a normalized, weighted measure of sentiment, ranging from -1 (most extreme negative) to 1 (most extreme positive).

For a deeper emotional analysis, raw emotion scores are evaluated based on the NRC Emotion Lexicon, which associates words with eight primary emotion categories—anger, anticipation, disgust, fear, joy, sadness, surprise, and trust—as well as two main sentiments (positive and negative). While the raw emotion scores offer a quick snapshot of the number of words associated with each emotion or sentiment, this counting method does not account for the number of emotion-related words in the text or a word's potential link to multiple emotion categories. Therefore, a weighted approach is taken, where emotion scores for each unique word are calculated based on their emotional associations and occurrence in the text. These scores are then summed across categories to generate final, normalized emotion scores that add up to 1. All analyses are replicated over time and for each debate topic.

The debate transcript is broken down into different sections by topic, marked by the hosts' announcement each time a new topic is introduced or when the time for a topic is exhausted. The

seven topics covered during the debate, in chronological order are: economy, abortion rights, immigration issues, transition of power and the election, foreign policy (including the Israel-Hamas war, Russian-Ukraine war, and the situation in Afghanistan), healthcare, and climate change. The closing statements are also analyzed to see how the two candidates emerged at the end of the debate.

4. Results and Discussion

4.1.Overall sentiment

Overall, both Trump and Harris displayed a positive polarity score, 0.033 and 0.051 respectively, indicating the sentiment conveyed by both during the debate is slightly more positive than negative. In terms of subjectivity, Trump scored 0.507, which is noticeably higher than Harris' 0.394. Analyzing the strength of the sentiments in Table 1, it is observed that the proportion of positive speech from Harris is almost double that of her negative speech. In contrast, Trump's proportions are relatively balanced. While the Democratic representative's overall sentiment leans more towards the positive side, the Republican counterpart's sentiment is somewhat mixed between positive and negative. Figure 1 shows the normalized emotion scores, in which the ranking of eight emotions are the same for both candidates, with ‘trust’ being the highest score, followed by ‘anticipation’ and ‘fear’.

4.2.Sentiment and emotion over time

The cumulative distributions of polarity and subjectivity measured over each candidate's speech are shown in Figure 2. Although the Kolmogorov-Smirnov tests were not statistically significant at a confidence level of 0.05, observation still shows that 80 percent of Harris’s subjective scores distribution is lower than 0.5, compared to only 60 percent for Trump.

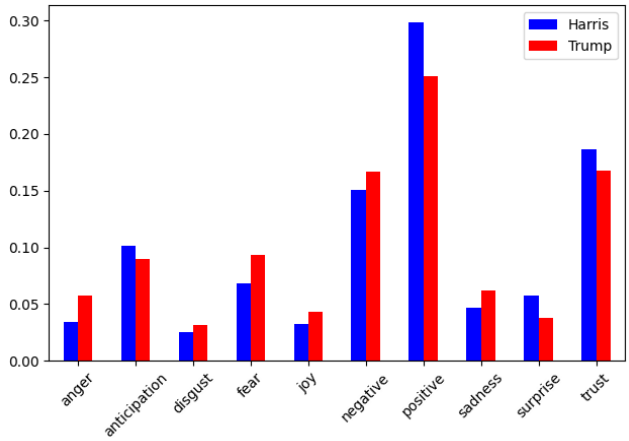

Fig 1. Emotion scores measured over the entire debate normalized on a scale of [0, 1]

	Positive	**Neutral**	**Negative**	**Compound**
Harris	0.223	0.654	0.122	0.9998
Trump	0.162	0.665	0.173	0.9992

Table 1. Sentiment strength measured over the entire debate

The cumulative distributions of sentiment intensity (measured at individual speech level) (Figure 3) shows a statistically significant difference in cumulative distributions of positive scores and compound scores between the two candidates. In general, the results so far have all shown that Harris conveys more positive sentiment over the course of the debate than Trump did. Other than 'joy', 'disgust', and 'sadness', the cumulative distribution of other emotions (Figure 4) between the two candidates are roughly similar.

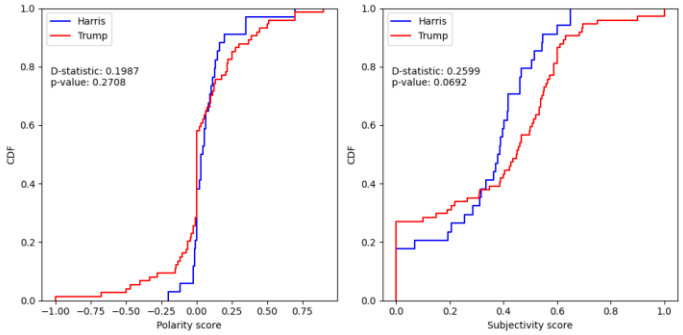

Fig 2. Polarity and subjectivity by speech – cumulative distribution

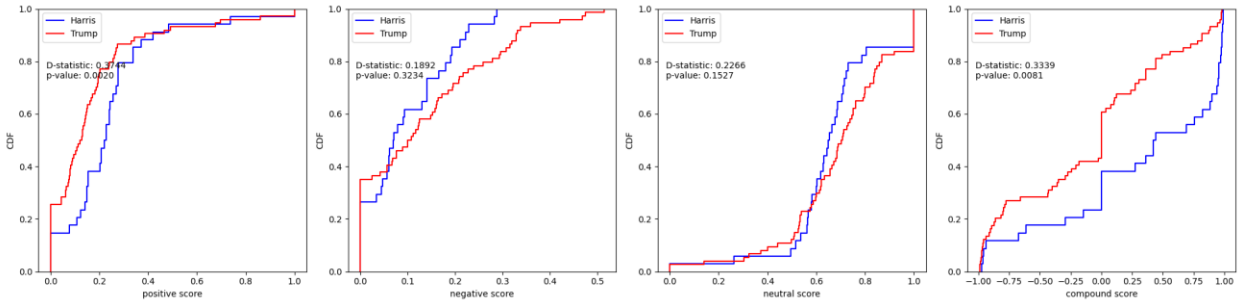

Fig 3. Sentiment intensity by speech – cumulative distribution

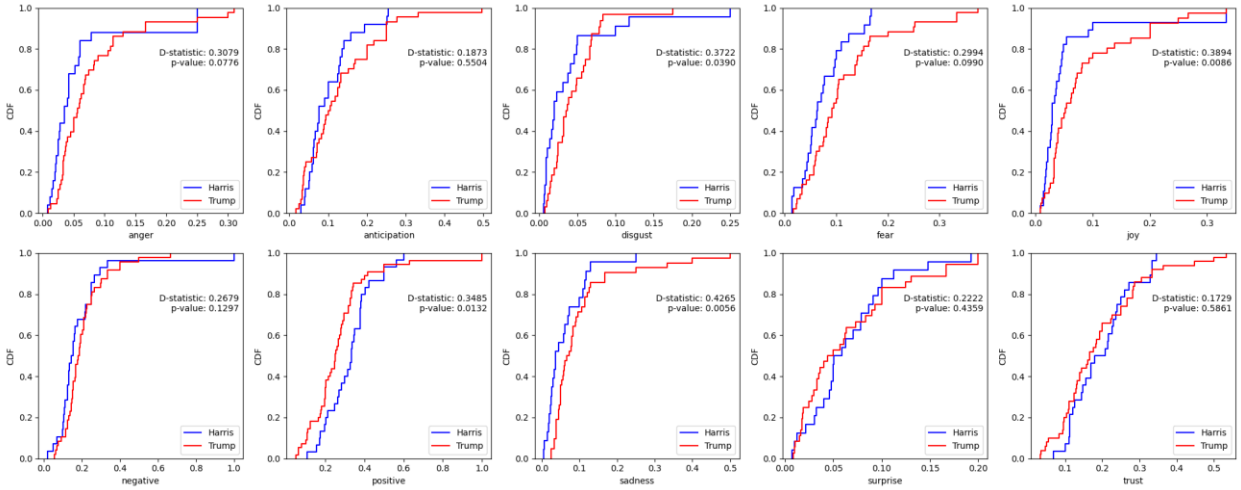

Fig 4. Emotion scores by speech – cumulative distribution

4.3.Sentiment analysis by topic

Across the four key topics of the debate, namely economy, abortion rights, immigration, and foreign policy, Harris was consistently more objective than Trump. Sentiment-wise, the most observable difference is the opposite signs in compound scores for immigration and foreign policy issues. Trump's overall sentiment on these two issues was negative, highlighted by higher scores across emotions categories like 'anger', 'fear', and 'sadness', Harris, on the other hand, conveyed more 'trust' and 'anticipation', which contribute to her more positive sentiment.

For economy and abortion, although the initial assessment of sentiment strength for both Harris and Trump indicates higher positivity, the subsequent breakdown of emotion scores reveals several key differences in which particular emotions evoked from each candidate. During

the debate on economic issues, Trump displayed a higher score for ‘trust’, but also for ‘anger’ and ‘fear’, whereas Harris conveyed noticeably more ‘disgust’, ‘fear’, and ‘sadness when the topic of abortion was introduced. These differences altogether shed more light on the differences in the two representatives' stances towards these topics as well as the focus of their presidential campaigns. Trump appeals to voters through his focus on issues pertaining to the economy, and foreign policy, whereas Harris’ campaign bolsters on abortion rights and racial justice.

Topic	harris_polarity	trump_polarity	harris_subjectivity	trump_subjectivity
economy	0.0015	0.0510	0.3213	0.5375
abortion	0.0687	0.1462	0.3936	0.5213
immigration	0.0202	0.0851	0.3865	0.4702
transition_of_power	0.0197	-0.0575	0.3920	0.5468
foreign_policy	0.0768	-0.0469	0.4433	0.5053
healthcare	0.0844	0.1427	0.3978	0.4749
climate	0.1967	-0.1017	0.4143	0.6100
closing_statement	0.0336	0.0776	0.3724	0.4276

Table 2. Polarity and subjectivity scores measured by topic and closing statements

	harris_neg	trump_neg	harris_neu	trump_neu	harris_pos	trump_pos	harris_ compound	trump_ compound
economy	0.1240	0.1130	0.6920	0.7100	0.1840	0.1770	0.9812	0.9849
abortion	0.1430	0.1050	0.6580	0.7660	0.1990	0.1290	0.9073	0.8578
immigration	0.1600	0.2020	0.6480	0.6320	0.1920	0.1650	0.8915	-0.9959
transition_of_power	0.2110	0.1700	0.6560	0.7160	0.1330	0.1140	-0.9660	-0.9898
foreign_policy	0.1280	0.2450	0.6420	0.5890	0.2310	0.1660	0.9978	-0.9989
healthcare	0.0930	0.1240	0.6500	0.6760	0.2570	0.2000	0.9924	0.9603
climate	0.0790	0.0550	0.6450	0.7850	0.2760	0.1600	0.9775	0.9105
closing_statement	0.0450	0.1850	0.6770	0.6200	0.2780	0.1960	0.9882	-0.3506

Table 3. Sentiment intensity measured by topic and closing statements

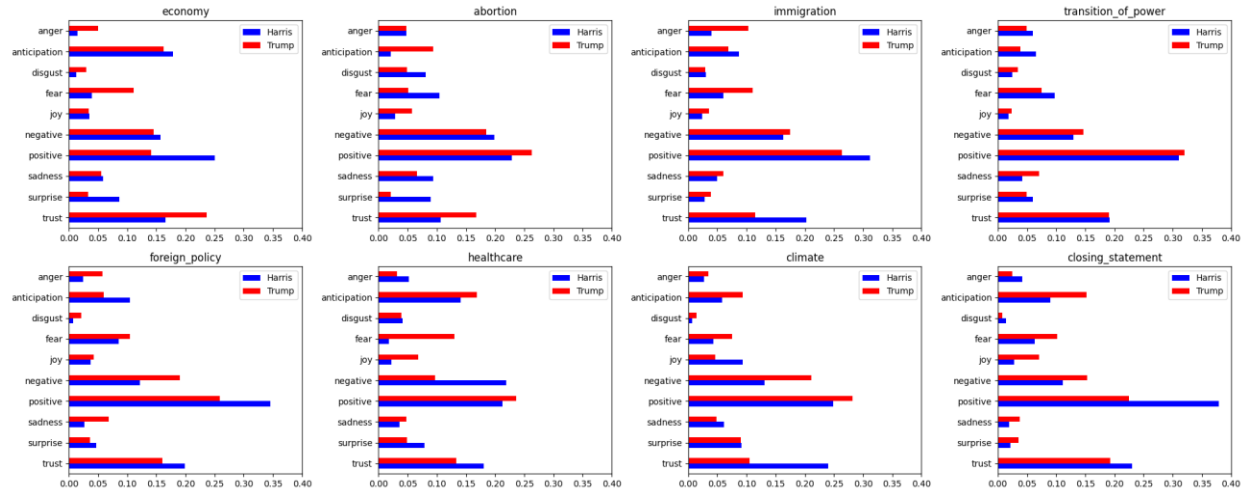

Fig 5. Emotion scores measured by topic and closing statements

Wrapping up the debate, Harris once again displayed a significantly more positive tone in her closing remarks. As for Trump, the closing statement's sentiment was still leaning more towards negative, albeit much less remarkable compared to during the debate.

5. Conclusion

The sentiment analysis conducted on the transcript of the 2024 U.S. Presidential Debate on September 10 has uncovered the key differences in the sentiment conveyed by Vice President Kamala Harris and Former President Donald Trump on several key issues, thereby further pinpoint the thematic ideologies of their campaigns and lobbying strategies. The results showed higher and more consistency in positive sentiment and overall objectivity from Harris. On the contrary, measures for Trump suggest much higher volatility in his sentiment and emotion throughout the debate. Despite the distinguishing, if not opposing, take on key issues, at the end of the day, trust almost always emerges among the top emotions over the course of the entire debate for both candidates. While this analysis is not meant to be a conclusive, in-depth political

analysis of the election or the campaigns of the two parties' representatives, the results can certainly add to the context of future attempts to dissect the race for the White House.

References

Bailey, M. M. (2019). *NRCLex* (Version 3.0.0) [Computer software]. Python Package Index. https://pypi.org/project/NRCLex/

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Toolkit: Sentiment analysis module* [Computer software]. nltk.org. https://www.nltk.org/api/nltk.sentiment.html

Hoffman, R. (2024, September 11). *READ: Harris-Trump presidential debate transcript*. ABC News. https://abcnews.go.com/Politics/harris-trump-presidential-debate-transcript/story?id=113560542

Loria, S. (n.d.). *TextBlob: Simplified text processing* (Version 0.18.0) [Computer software]. TextBlob Documentation. https://textblob.readthedocs.io/en/dev/

Mohammad, S. (n.d.). *NRC emotion lexicon*. National Research Council Canada. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm