Automated Political Affiliation Ranking via Subjectivity Detection in Political Discourse

IST 664 – Natural Language Processing

David Caspers

Objectives

1

Develop a model to classify statements as **objective** or **subjective.**

2

Extend the model to classify public statements from members of the **US House and Senate** as objective or subjective.

3

Implement a methodology to **rank politicians** on a political spectrum based on their opinion statements.

Why Is this Interesting?

REDUCE BIAS AND EFFORT

IMPROVE TRANSPARENCY

TRACK CHANGES OVER TIME

About the Data

Labeled Subjectivity Corpus

Source: NewsSD-ENG (English news articles).

Topics Covered: Law, civil rights, economics, and other controversial political subjects.

Labels:

- **Subjective:** Based on or influenced by personal feelings, tastes, or opinions.
- **Objective:** Factual statements, free from personal influence.

Usage: Train and evaluate sentence classifier

VoteSmart API

Source: Non-partisan organization.

Data: Public statements (Tweets, interviews, speeches) from **2022-2023** for all then-elected members of Congress

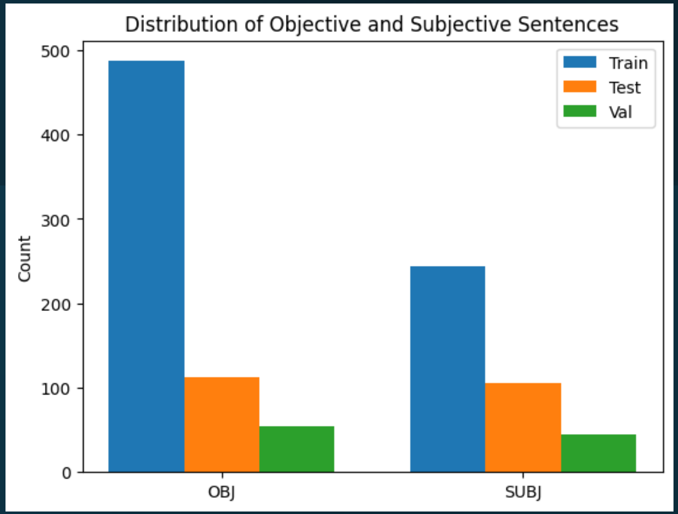
Usage: corpus of politician's publicly stated political opinions

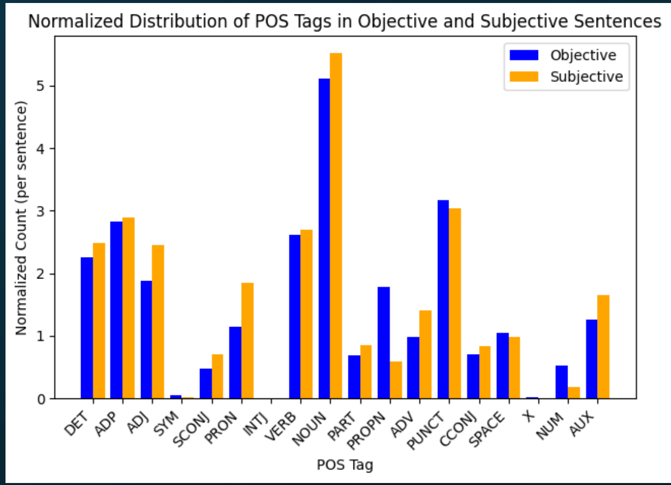
Limited Government Scorecard, Institute for Legislative Analysis

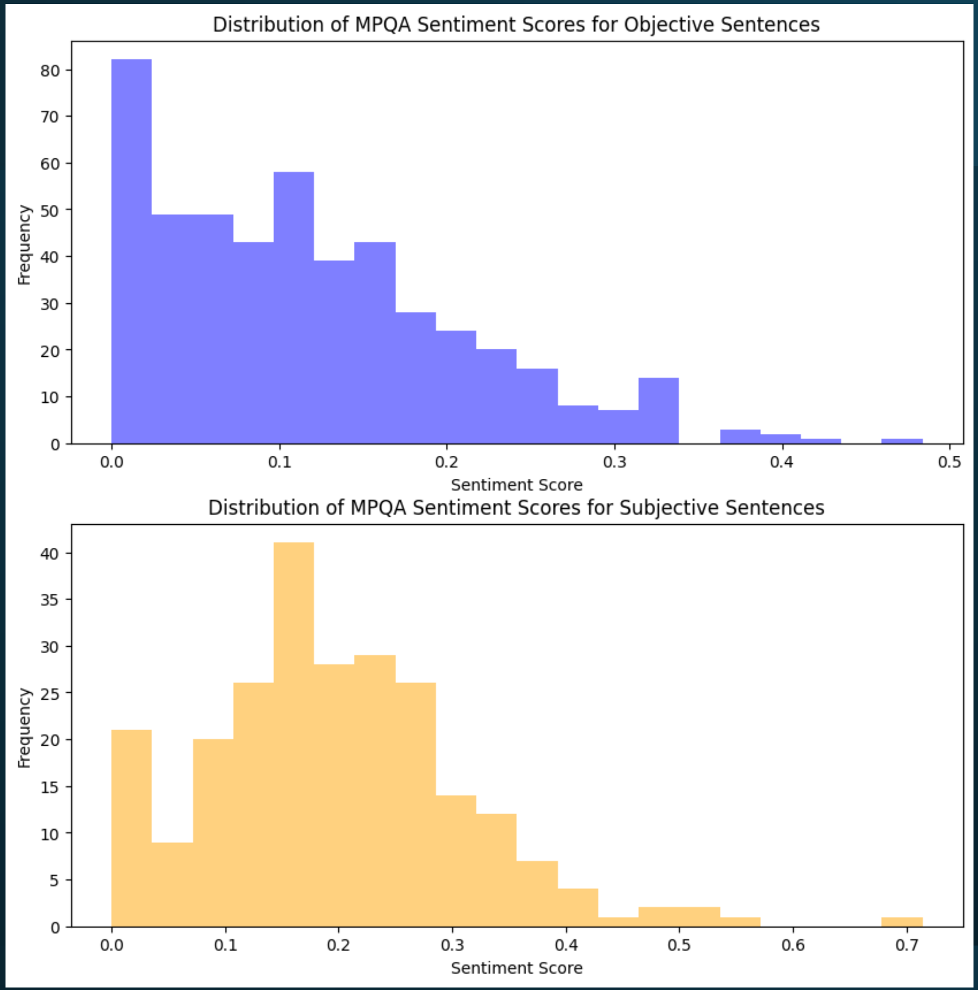
Purpose: Ranks politicians based on their voting records reflecting commitment to **limited government principles**.

Usage: Benchmark to compare my methodology's rankings with expert opinions.

Exploratory Analysis







Results

Model ***all Naïve Bayes models use TFIDF***	**Accuracy** **BNB / MNB**	**Precision** **BNB / MNB**	**Recall** **BNB / MNB**	**F-1** **BNB / MNB**
Bag of Words (Unigram)	53.42% 49.31%	62.83% 75.24%	53.42% 49.32%	44.95% 33.56%
Bigrams + Trigrams	47.95% 47.95%	43.94% 46.21%	47.95% 47.94%	32.90% 34.32%
Unigram + Bigram + Trigram	**65.66%** **67.44%**	**64.98%** **64.14%**	**65.66%** **67.44%**	**65.81%** **62.72%**
Part of Speech Appended to Words, Unigram, Bigram, Trigram	50.22% 49.77%	54.32% 75.35%	50.22% 49.77%	40.93% 34.53%
Unigram, Bigram, Trigram, and POS Counts	64.84% 48.86%	69.45% 57.92%	64.84% 48.86%	63.12% 33.35%
Unigrams + Bigrams + Trigrams + MPQA Sentiment Score	65.30% 49.32%	69.81% 62.34%	65.30% 49.32%	63.68% 34.31%
Tuned "distilbert-base-uncased"	**75.80%**	**76.73%**	**75.80%**	**75.68%**

BERT Classifier Model Applied to Political Discourse

Text: Today Rep. Ilhan Omar released the following statement to commemorate the one year anniversary of the January 6th insurrection.

Predicted Class: OBJ

Class Probabilities: [0.9988122 0.00118786]

Text: "Today marks one year since the attacks of January 6th.

Predicted Class: OBJ

Class Probabilities: [9.990722e-01 9.278190e-04]

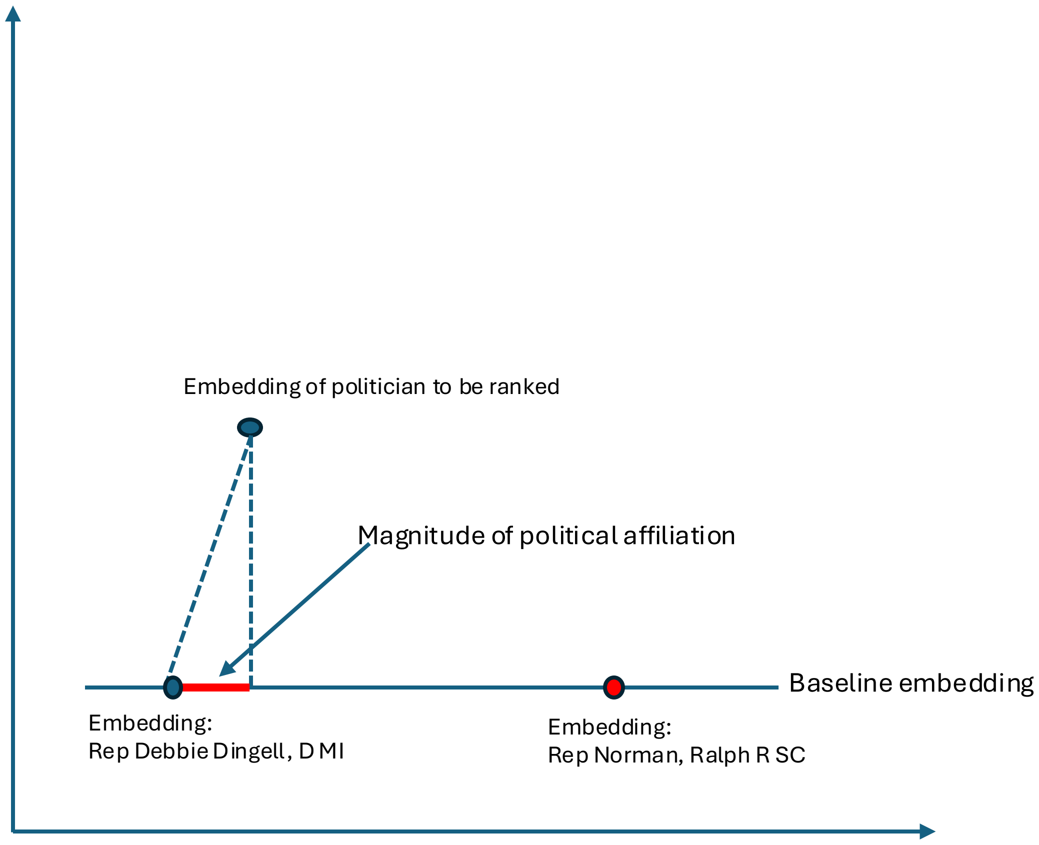
Text: I will never forget the experience of fearing for my life, my fellow members, and staff on a day designed to show the strength of our democracy.

Predicted Class: SUBJ

Class Probabilities: [0.00251374 0.99748623]

Methodology for Inferring Political Alignment

1. Classify each sentence of Politicians Public Statements
2. Only retain Subjective Sentences
3. Embed Sentences Using Pre-trained Bert Sentence Embedder “all-MiniLM-L6-v2”
4. Average each politician’s embeddings to get one ‘representative’ embedding
5. Use Politicians on opposite side of political spectrum (as defined by the Limited Government Scorecard) as base case and project embeddings onto vector
6. Rank by Magnitude
7. Compare to Limited Government Scorecard



Projecting Politicians Political Spectrum

Conclusions:

- Delineation between subjective and objective content can be unclear
- Quality and quantity of available data matters

– Methodology seems promising though requires significant tuning and cleaner/higher quality data sources.

Politician	The Institute for Legal Analysis Ranking*	Predicted Ranking	Sentences Available To Draw Inferences From
Rep Chip Roy	1	45	48
Rep Andy Biggs	2	24	75
Rep Matt Rosendale	3	70	9
Rep Lauren Boebert	4	12	167
Rep Greg Steube	5	109	109

* Reps with no available public statements omitted from ranking

Questions?