Improving Measures of Emotion in Judicial Text using Word Embeddings

Peter Bachman [1]

[1]Washington University in St. Louis

Literature Review

- Text analysis is frequently used in judicial politics
- Recent work uses dictionary-based methods to measure emotion and sentiment (Ballingrud 2021; Budziak, Hitt, and Lempert 2019; Owens and Wedeking 2011)
- Legalese is technical and dictionaries may miss nuance in text
- Can unsupervised word embedding models combined with dictionaries improve measures of emotion in legal text?

Methods and Data

- I calculated the average similarity of words in the legal text to words in the positive and negative sentiment dictionaries as follows:

$$\frac{\sum_{i=1}^{n} A_{\cos\ i}}{n} - \frac{\sum_{j=1}^{m} B_{\cos\ j}}{m}$$

- $A_{\cos\ i}$ = Cosine similarity between corpus word and positive sentiment word i
- $B_{\cos\ j}$ = Cosine similarity between corpus word and negative sentiment word j

Data

- Pulled Supreme Court opinion text from 2015 to 2019 using Caselaw Access Project's API
  - Docket numbers provided by the Supreme Court Database
- This corpus was tokenized, lowercased, had punctuation removed, and had stop words removed in that order
  - Resulting corpus had 1,398,758 total words/tokens and 33,679 unique words
- GloVe Word Embedding Model using Wikipedia 2014 + Gigaword 5 pre-trained word vector
- The Afinn and Bing Sentiment Lexicons for measures of positive and negative sentiment

Results

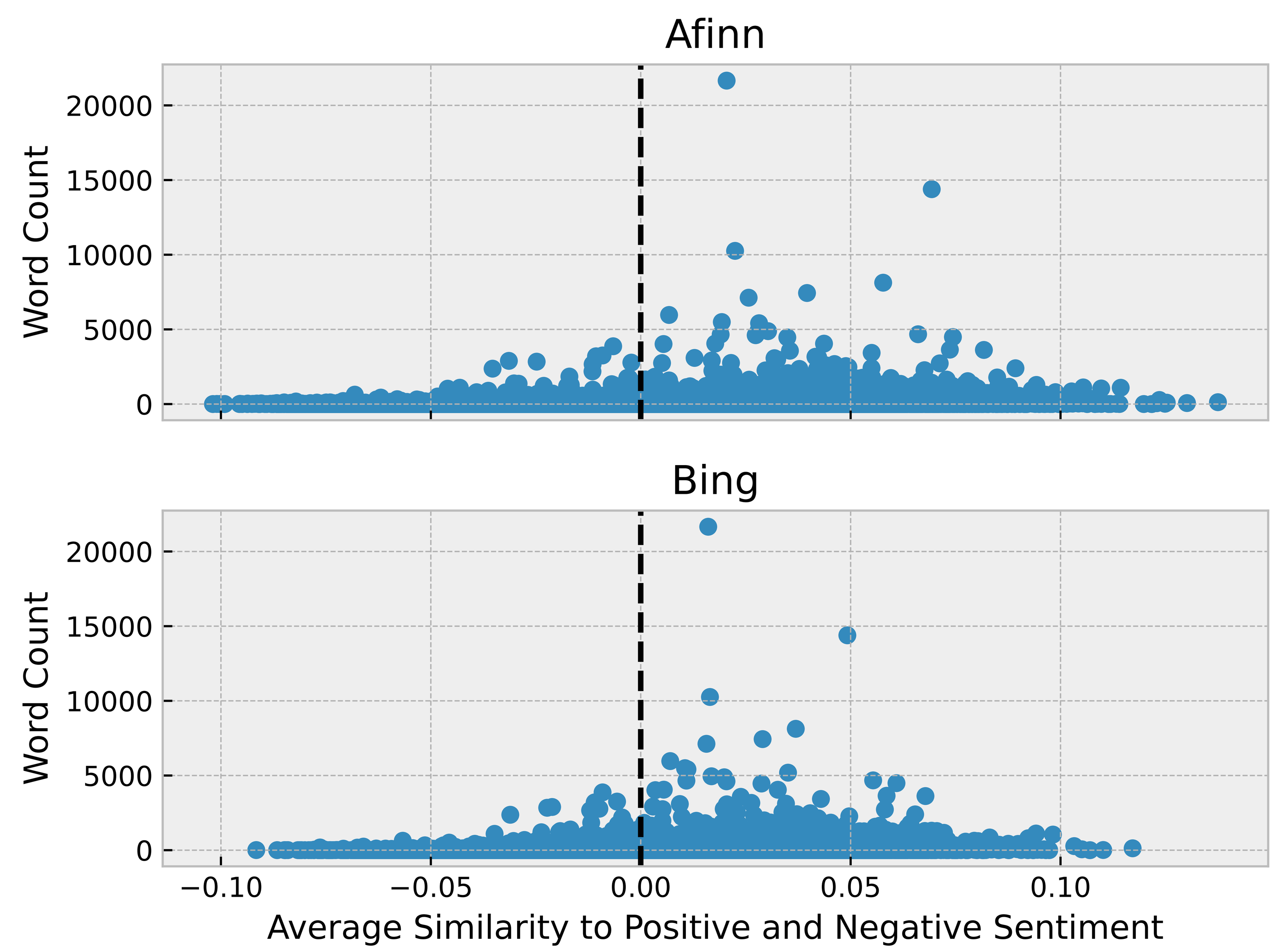
Word	Average Similarity	Word Count
unique	0.117179	126
thanks	0.110177	12
showcase	0.107021	1
achieved	0.105090	58
offers	0.103266	273
⋮	⋮	⋮
mistreatment	-0.084070	5
suspecting	-0.084646	4
mismanagement	-0.084847	5
blaming	-0.086605	4
exacerbated	-0.091583	8

Table 1. Most Positive and Negative Words in Corpus According to Bing Sentiment Similarity

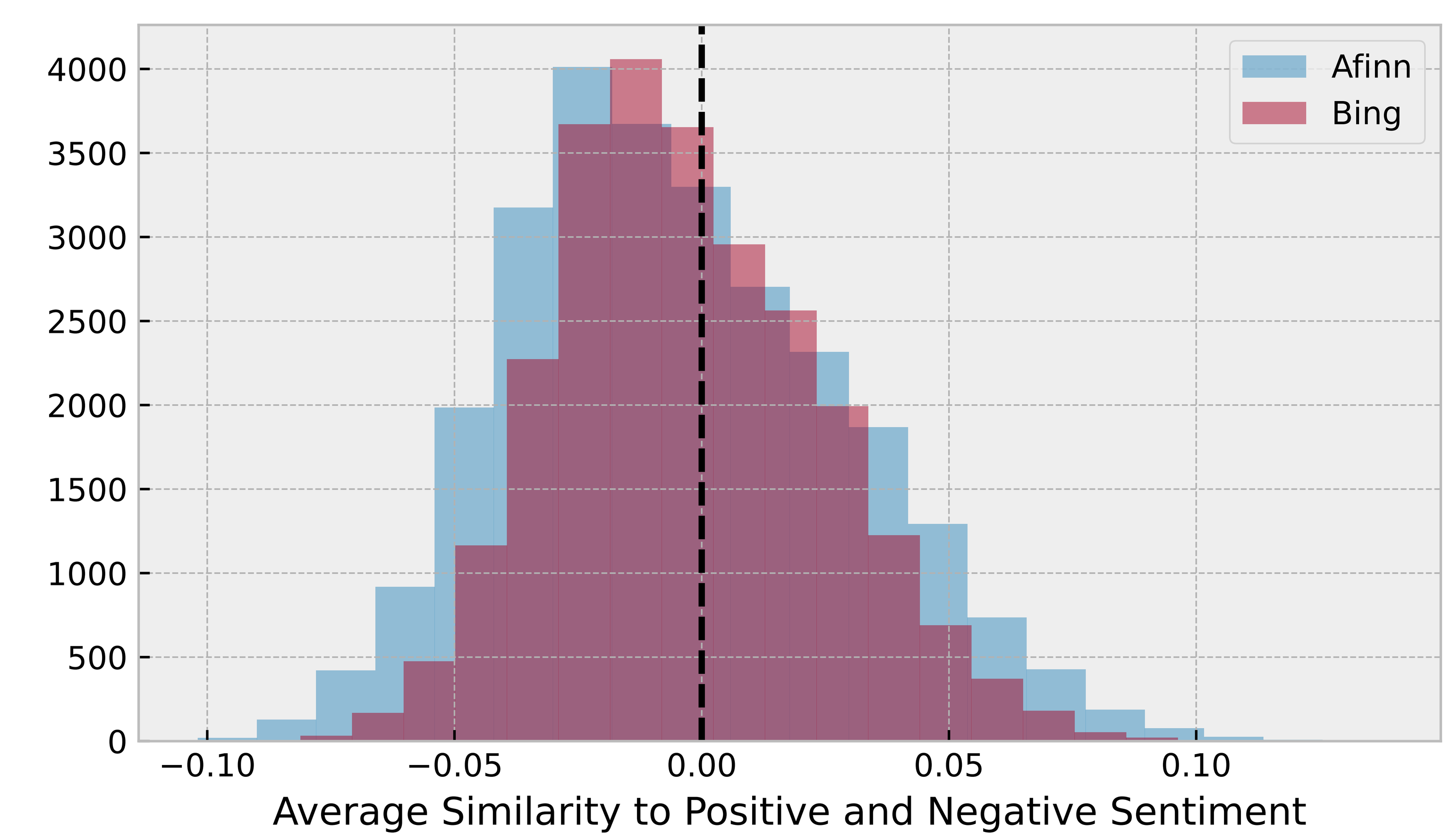
Word	Average Similarity	Word Count
unique	0.137496	126
enjoyed	0.130125	68
contribution	0.125302	93
partnership	0.124888	22
offers	0.123550	273
⋮	⋮	⋮
mismanagement	-0.099101	5
castigated	-0.099141	1
mistreatment	-0.100295	5
exacerbated	-0.100980	8
debilitated	-0.101905	2

Table 2. Most Positive and Negative Words in Corpus According to Afinn Sentiment Similarity

Average Sentiment Similarity and Word Count

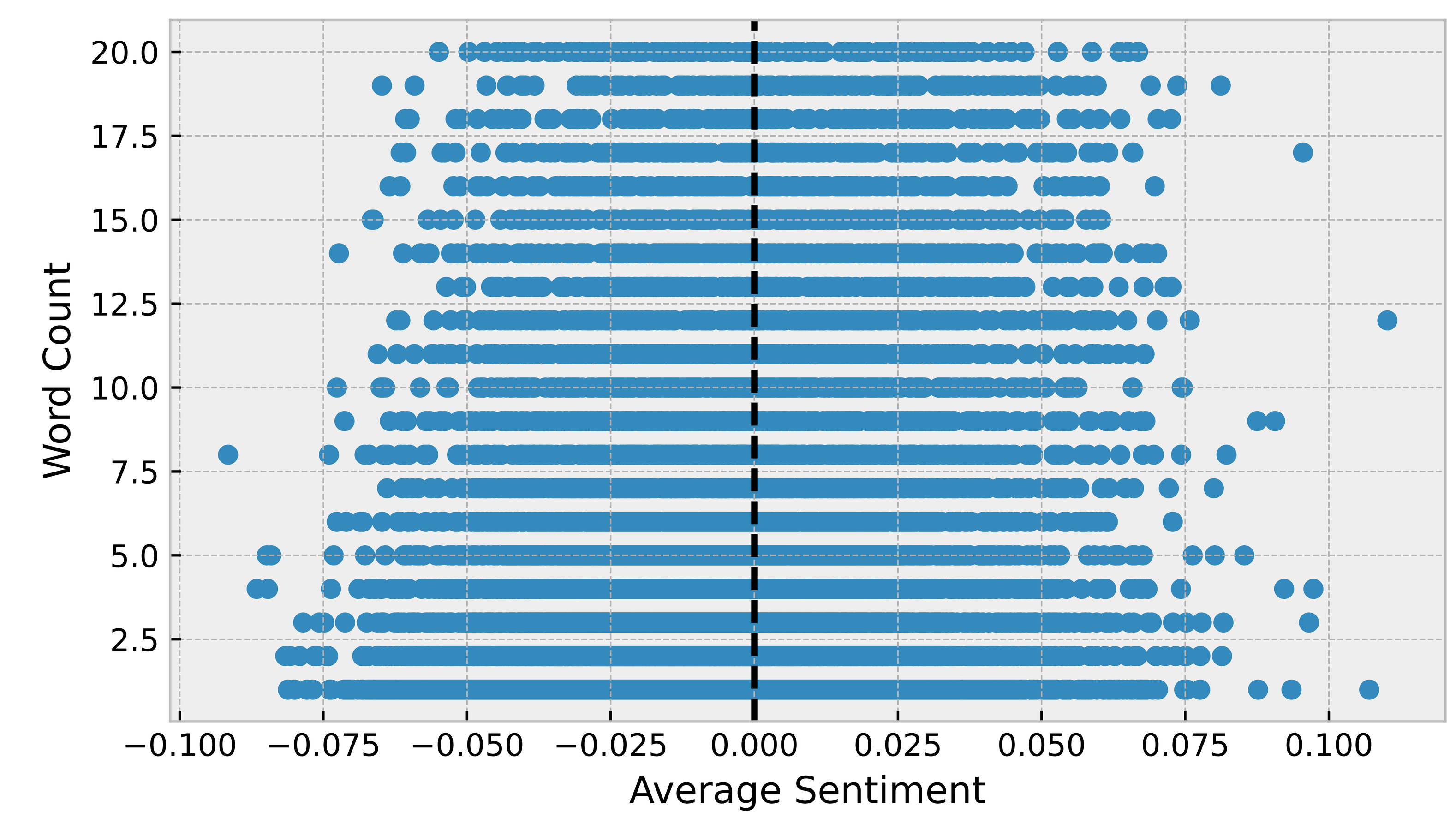


Average Sentiment Similarity Distribution



Top 100 Positive (Red) and Negative (Blue) Words

Average Sentiment Similarity of Words occuring 20 Times or Less (Afinn)



Current Conclusions

- A few words that have some measure of positive or negative sentiment, but the range of average sentiment is small.
- The results are inconclusive at this point

Next Steps

1. Use these findings as the basis for my NSF GRFP application.
2. Use more recent forms of word embeddings (e.g. GPT-3, BeRT) instead of GloVe.
3. Train a word embeddings algorithm to legal text instead of using an "off the shelf" word embedding vector
4. Incorporate weights used by some sentiment dictionaries to properly average positive/negative measure
5. Use LWIC as a direct comparison to prior judicial research
6. Use lexicons with more than two dimensions of emotion

References

[1] Caselaw Access Project.

[2] Gordon Ballingrud. Ideology and Risk Focus: Conservatism and Opinion-Writing In the U.S. Supreme Court. 102(1):281–300.

[3] Jeffrey Budziak, Matthew P. Hitt, and Daniel Lempert. Determinants of Writing Style on the United States Circuit Courts of Appeals. 7(1):1–28.

[4] Justin Grimmer. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton.

[5] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, page 168. ACM Press.

[6] Finn Årup Nielsen. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.

[7] Ryan J. Owens and Justin P. Wedeking. Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions. 45(4):1027–1061.

[8] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

[9] Harold J. Spaeth, Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. Supreme Court Database, Version 2021 Release 1.

QR Code to Replication Repository