Automated Classification of Political Video Without Text

Christopher Lucas
Department of Political Science

Ishita Gopal
Transdisciplinary Institute in Applied Data Sciences

Washington University in St. Louis
Arts & Sciences

Can audio features alone identify attack advertisements in the US?

- Political campaigns heavily rely on audiovisual communication.
- Audio data = speech + background music
- Happy/upbeat music → positive ad
- Sad/ominous music → attack ad

Model accurately classifies using audio signals alone, without text.

Wav2vec2 Model

- For speech classification tasks
- Learn features (pretrain) from the raw audio
- Requires relatively little labeled data

Fine Tune Wav2vec2 on televised ads collected by WMP

- 2018 midterms which include Senate, House, Governor, Downballot campaign ads.
- Labels: "Negative", "Positive", or "Contrast" (combine contrast and positive ads)
- Split ~10,000 ads into 80%, 10%, 10% train/validate/test splits
- Balanced accuracy ~ 84%

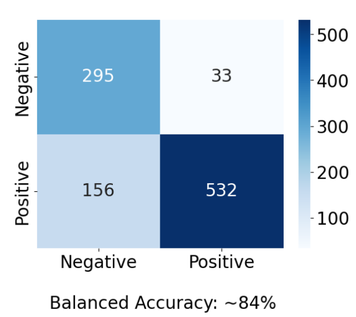


Fig 1: Prediction on WMP test set (Contrast = Positive)

Negative ads more likely to appear on Google with distribution driven by party

Fig 2: Top 20 advertisers on Google

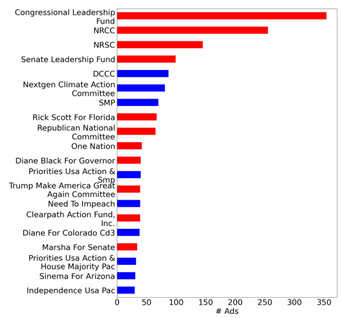
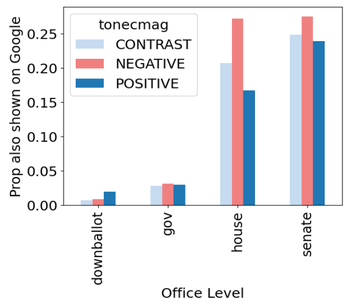


Fig 3: Proportion of WMP ads found on Google by office



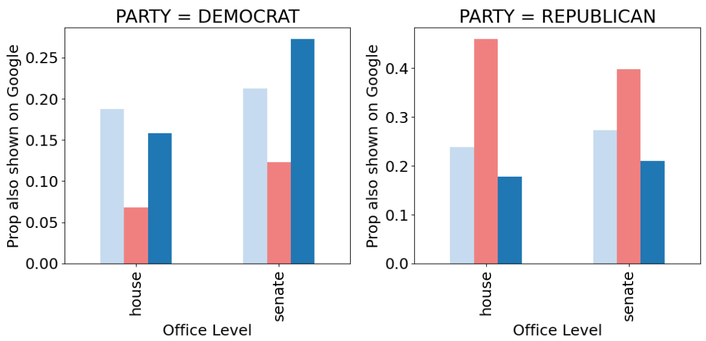


Fig 4: Proportion of WMP ads found on Google by party

Out of Domain Performance on Google Ads

Hand code 100 as a validation check

- Balanced accuracy ~72%
- High Precision in identifying negative class ~90%

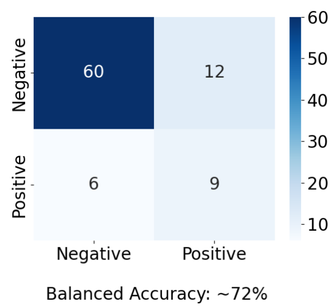


Fig 5: Out of domain prediction on hand labeled Google Ads

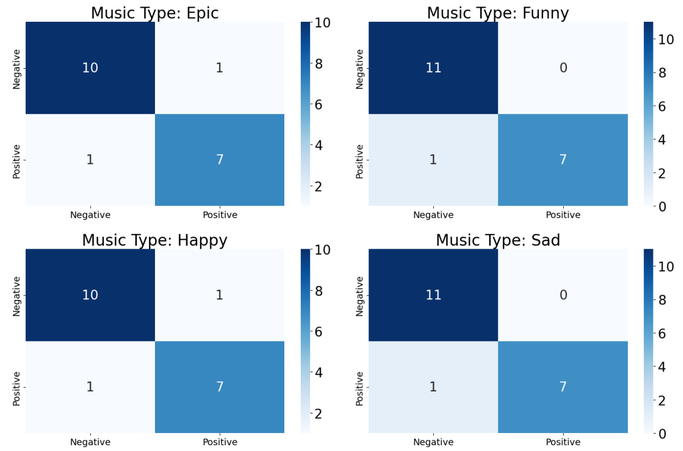
Label 3200 ads shown in the runup to 2018 election

- ~51% ads classified as negative
- Opposite of what was found on Facebook

Background music does not drive predictive accuracy

- Extract vocals from 19 ads in the WMP dataset
- Combine vocals with 4 types of background music (Happy, Sad, Funny, Epic)
- 19 x 4 = 76 ads
- Predict on simulated data

Fig 5: Prediction across 4 background music types



Next Steps

- Improve out of domain classification performance
- Augment training data with google ads data
- Link google data with candidates
- Fine-tune for different countries/languages and test the extrapolation of these language-agnostic models to identify attack ads cross-nationally

References

Tarr, Alexander, June Hwang, and Kosuke Imai. "Automated coding of political campaign advertisement videos: An empirical validation study." *Political Analysis* 31.4 (2023): 554-574.

Fowler, Erika Franklin, Michael Franz, and Travis Ridout. *Political advertising in the United States*. Routledge, 2021.

Fowler, Erika Franklin, et al. "Political advertising online and offline." *American Political Science Review* 115.1 (2021): 130-149.

PolMeth 2024, UC Riverside