PolitiGauge

Extreme | Left | Left-Center | Least Biased | Right-Center | Right | Extreme

Matt, Taber, Yoseph

Problem

- *What we are trying to do:*
    - Battle the growing problem of filter bubbles to promote a more informed public
- *Why it is hard:*
    - Number of different features
    - Subtlety
- *Objectives*:
    - Create a high accuracy political bias predictor
    - Incorporate it in an intuitive chrome extension

Why is it Important?

Filter bubble: *a phenomenon in which a person is exposed to ideas, people, facts, or news that adhere to or are consistent with a particular political or social ideology.*

- *Who cares?*
    - The entirety of news-consuming Americans.
- *What difference it could make?*
    - Helping American news consumers become self-aware
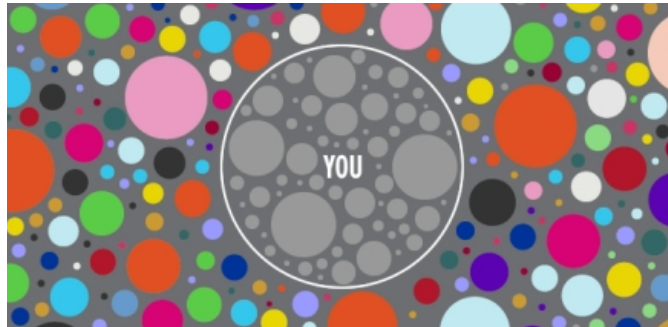
Current Landscape [Lit. Survey]

How it is done today:

- Classifiers, not regression models [1]
- Crowdsourced labeling [2]
- NLP of *articles* only [3]

Limitations:

- Trustworthiness of crowdsourced data
- Lack of multi-feature training

How We're Different

1) Chrome extension
2) Regression instead of classifier
   a) Implemented with a combination of best features from our lit. review
3) Taking into account:
   a) Individual articles [AllSides.com]
   b) Political speeches [Convote and Congressional text dataset]
   c) Overall organizational bias [MediaBiasFactCheck and AllSides.com]

Risk/Cost/Time

Challenges and Risks:

- Determining our baseline for our political spectrum
- Building regression model(s)
- Incorrectly gauging bias for our future users

Payoffs:

- Chrome extension helping users break out of echo chambers
- Convince authors to write their articles in a "non-partisan" way

Measuring Success

Mid-term:

- Have our initial datasets organized and cleaned
- Basic model(s) selected.

Final:

- Trained and cross-tested multiple models across multiple datasets
- Compare our model(s) with baseline models from literature
- Prototype of browser extension

Plan of Activities

1. Cleaning and Analyzing the datasets
   a. Convote, Congressional Record for the 43rd-114th Congresses
   b. Web-scraping and Allsides.com,
   c. Ideological Books Corpus
2. After Initial Data exploration
   a. Coming up with the "Gauge" of political bias
   b. Model selection
3. Model Training
4. Write-up/Prototyping

Questions?

References

Journal Articles

[1] - http://www.aclweb.org/anthology/P14-1105
[2] - https://www8.gsb.columbia.edu/media/sites/media/files/JustinRaoMediaBias.pdf
[3] - https://cs224d.stanford.edu/reports/MisraBasak.pdf

Pictures

Extension: https://www.found.co.uk/blog/5-awesome-browser-extensions/#.XG2U3OhKg2w
Gauge Pic: https://mediabiasfactcheck.com/fivethirtyeight/
Questions Pic: https://leadingwithtrust.com/2017/04/30/10-questions-great-bosses-regularly-ask-their-people/
Chrome Extension: https://www.pivotaltracker.com/integrations/chrome
Filter Bubble: https://www.redcaffeine.com/red-talks/posts/the-social-media-filter-bubble/
Database pic: http://www.robots.ox.ac.uk/~vgg/data/lip_reading/
All Sides pic: https://www.allsides.com/media-bias/media-bias-ratings