Text-Based Ideal Points

Keyon Vafa
Columbia University

Joint work with:

Suresh Naidu
Columbia University

David Blei
Columbia University
Ideal Points

- Red line: Nominated by a Republican
- Blue line: Nominated by a Democrat

Source: Lee Epstein, Washington University in St. Louis; and Andrew D. Martin and Kevin Quinn, University of Michigan

Image Source: New York Times
Ideal Points

Bayesian Ideal Points

- Probabilistic method to measure political positions of legislators
- Based solely on voting record

\[ v_{ij} \sim \text{Bern}(\sigma(\beta_j + x_i \eta_j)) \]
Vote Ideal Points

Analyze votes on shared bills to infer political positions.

Limitations:

- Cannot compare groups who do not vote together (e.g. judges on different courts).

- Votes on decisions must be available (e.g. cannot extend to presidential candidates).

Solution: Text-based ideal points!

- Analyze language of speeches to infer political preferences.
Vote-Based Ideal Points

IN: Voting Record

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susan Collins</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Elizabeth Warren</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>John McCain</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chuck Schumer</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

OUT: Ideal Points

John McCain
Susan Collins
Chuck Schumer
Elizabeth Warren

IN: Voting Record
OUT: Ideal Points
Elizabeth Warren: Donald Trump spent years pedaling Trump University, a sham college that his own former employees shared.

Chuck Schumer: My final question is this: Since we have a Department of Homeland Security that needs funding, and the issue of budget for the homeland security.

Susan Collins: I wish to commemorate the 200th anniversary of the Town of Woodstock. Known today as a gateway to the

John McCain: I would like to thank my friend and colleague from Indiana for his “Waste of the Week” speech, although it was

IN: Speeches

OUT: Ideal Points + Ideological Topics
Existing Methods

Existing methods for inferring political positions from text either:

- Use party labels
- Combine text with voting records
- Use hand-labeled political text
- Require grouping of texts into single issues
Text-Based Ideal Points

The Text-Based Ideal Point Model (TBIP) is completely unsupervised:

- Does not require party labels, voting records, hand-labeled political text, or grouping of text into single issues

Advantages of being unsupervised:

- Applicable to unlabeled political discourse
- Does not force hard membership into binary groups
- Does not depend on subjectivity of coders
Political Framing

Entman’s definition of framing (Entman, 1993):

“[Selecting] some aspects of a perceived reality and [making] them more salient in a communicating text, in such a way as to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.”

Political framing: When discussing a topic, word choice is affected by political message.

Frames for abortion (Boydstun et al., 2014; Johnson et al., 2017):

- “life” and “unborn” invoke morality and religion
- “choice” and “freedom” invoke constitutionality and personal liberty
Text-Based Ideal Points

Vote-based ideal points:

- Inferred by vote differences on shared bills.

Text-based ideal points:

- Inferred by word choice differences on shared topics.
Model

The TBIP is based on **Poisson factorization**:

\[ y_{dv} \sim \text{Pois} \left( \sum_k \theta_{dk} \beta_{kv} \right) \]

- **word counts**
- **document intensities**
- **topics**

We add two terms to the Poisson factorization log-likelihood:

\[ y_{dv} \sim \text{Pois} \left( \sum_k \theta_{dk} \beta_{kv} \exp\{x_{ad} \eta_{kv}\} \right) \]

- **“ideological” topics**
- **ideal point for author of document** \( d \)
Inference

Posterior distribution for latent parameters \((\theta, \beta, \eta, x)\) is approximated with variational inference.

TensorFlow and PyTorch implementations are available at:

\[ \text{github.com/keyonvafa/tbip} \]
U.S. Senate Speeches
Ideal Points

Bernie Sanders (I-VT)
Elizabeth Warren (D-MA)
Sherrod Brown (D-OH)
Chuck Schumer (D-NY)
Amy Klobuchar (D-MN)
Susan Collins (R-ME)
Mark Warner (D-VA)
Rand Paul (R-KY)
Jeff Sessions (R-AL)
John McCain (R-AZ)
Ben Sasse (R-NE)
Marco Rubio (R-FL)
Mitch McConnell (R-KY)
U.S. Senator Tweets

209,779 tweets from senators between 2015-2017
Votes vs Speeches vs Tweets

Correlation to vote ideal points:
- Votes vs Speeches: 0.88
- Speeches vs Tweets: 0.94
2020 Democratic Presidential Candidate Tweets

45,927 tweets from 19 candidates between 2019-2020
2020 Democratic Candidates

#medicareforall, insurance companies, profit, health care
health care, plan, medicare, americans, care, access
healthcare, universal healthcare, public option, plan

more progressive          green new deal, fossil fuel industry, fossil fuel, planet, pass

more moderate

climate change, climate, climate crises, plan, planet, crisis
solutions, technology, carbon tax, climate change, challenges
Comparisons

Other methods: Wordfish (Slapin and Proksch, 2008) and Wordshoal (Lauderdale and Herzog, 2016)

Evaluate each ideal point method by measuring correlation and rank correlation to vote ideal points.

<table>
<thead>
<tr>
<th></th>
<th>Speeches 111</th>
<th>Speeches 112</th>
<th>Speeches 113</th>
<th>Tweets 114</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORDFISH</td>
<td>0.47</td>
<td>0.45</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>WORDSHOAL</td>
<td>0.61</td>
<td>0.64</td>
<td>0.60</td>
<td>0.56</td>
</tr>
<tr>
<td>TBIP</td>
<td><strong>0.79</strong></td>
<td><strong>0.73</strong></td>
<td><strong>0.86</strong></td>
<td><strong>0.85</strong></td>
</tr>
</tbody>
</table>
Recap

We develop an unsupervised model to learn ideal points and ideological topics solely from text.

Text-based ideal points can be used to learn political preferences for non-voting entities (e.g. presidential candidates).

We use an efficient variational inference algorithm to apply the model to large datasets.

All code (including Tensorflow and PyTorch implementations) available at:

www.github.com/keyonvafa/tbip
Thank you!
References


• VoxGovFEDERAL (2020). U.S. senators tweets from the 114th Congress.