Measuring Ideological Proportions in Political Speeches

Yanchuan Sim† Brice Acree* Justin H. Gross* Noah A. Smith†

†Language Technologies Institute
Carnegie Mellon University

*Department of Political Science
University of North Carolina at Chapel Hill

Conference on Empirical Methods in Natural Language Processing
October 19, 2013
Seattle, USA
“Well, I think you hit a reset button for the fall campaign. Everything changes. It’s almost like an Etch-A-Sketch. You can kind of shake it up and restart all over again.”

— Eric Fehrnstrom, Spokesman for Presidential Candidate Mitt Romney, 2012
“Well, I think you hit a reset button for the fall campaign. Everything changes. It’s almost like an Etch-A-Sketch. You can kind of shake it up and restart all over again.”

— Eric Fehrnstrom, Spokesman for Presidential Candidate Mitt Romney, 2012
Motivation

- Successful primary candidates “move to the center” before a general election.
Motivation

- Successful primary candidates “move to the center” before a general election.
  - In primary elections, voters tend to be *ideologically concentrated*.
Motivation

- Successful primary candidates “move to the center” before a general election.
  - In primary elections, voters tend to be *ideologically concentrated*.
  - In general elections, voters are now more *widely dispersed across the ideological spectrum*. 
Motivation

- Successful primary candidates “move to the center” before a general election.
  - In primary elections, voters tend to be *ideologically concentrated*.
  - In general elections, voters are now more *widely dispersed across the ideological spectrum*.

Hypothesis

Obama, McCain, and Romney should use more “extreme” ideological rhetoric in the primaries than the general election.

Can we measure candidates’ *ideological positions* from their prose at different times?
Conceptualizing "ideology"

- Ideology as **categorical** (a typology)
  - Patterns of commonly held political **beliefs**, **core values**, and **preoccupations** represented by reference to recognizable prototypes.
  - *Left, Right, Center, Religious left, Progressive, Libertarian, Far right, etc...*
Conceptualizing “ideology”

- **Ideology as *categorical* (a typology)**
  - Patterns of commonly held political *beliefs*, *core values*, and *preoccupations* represented by reference to recognizable prototypes.
  - *Left, Right, Center, Religious left, Progressive, Libertarian, Far right, etc...*

- **Ideology as *hierarchical***
  - Left vs Right – internally diverse set of finer grained distinctions.
Ideology tree

- Progressive
- Religious Left
- Far Left
- Center Left
- Religious Right
- Center Right
- Libertarian
- Populist
- Far Right

Measuring Ideological Proportions in PoliticalSpeeches (EMNLP 2013)
Cue-lag representation of speeches

Instead of putting more limits on your earnings and your options, we need to place clear and firm limits on government spending. As a start, I will lower federal spending to 20 percent of GDP within four years’ time – down from the 24.3 percent today.

The President’s plan assumes an endless expansion of government, with costs rising and rising with the spread of Obamacare. I will halt the expansion of government, and repeal Obamacare.

Working together, we can save Social Security without making any changes in the system for people in or nearing retirement. We have two basic options for future retirees: a tax increase for high-income retirees, or a decrease in the benefit growth rate for high-income retirees. I favor the second option; it protects everyone in the system and it avoids higher taxes that will drag down the economy.

I have proposed a Medicare plan that improves the program, keeps it solvent, and slows the rate of growth in health care costs.

— Excerpt from speech by Romney on 5/25/12 in Des Moines, IA
Instead of putting more limits on your earnings and your options, we need to place clear and firm limits on government spending. As a start, I will lower federal spending to 20 percent of GDP within four years’ time – down from the 24.3 percent today.

The President’s plan assumes an endless expansion of government, with costs rising and rising with the spread of Obamacare. I will halt the expansion of government, and repeal Obamacare.

Working together, we can save Social Security without making any changes in the system for people in or nearing retirement. We have two basic options for future retirees: a tax increase for high-income retirees, or a decrease in the benefit growth rate for high-income retirees. I favor the second option; it protects everyone in the system and it avoids higher taxes that will drag down the economy.

I have proposed a Medicare plan that improves the program, keeps it solvent, and slows the rate of growth in health care costs.

— Excerpt from speech by Romney on 5/25/12 in Des Moines, IA
Cue-lag representation of speeches

Instead of putting more limits on your earnings and your options, we need to place clear and firm limits on government spending. As a start, I will lower federal spending to 20 percent of GDP within four years’ time – down from 24.3 percent today.

The President’s plan assumes an endless expansion of government, with costs rising and rising with the spread of Obamacare. I will halt the expansion of government, and repeal Obamacare.

Working together, we can save Social Security without making any changes in the system for people in or nearing retirement. We have two basic options for future retirees: a tax increase for high-income retirees, or a decrease in the benefit growth rate for high-income retirees. I favor the second option; it protects everyone in the system and it avoids higher taxes that will drag down the economy.

I have proposed a Medicare plan that improves the program, keeps it solvent, and slows the rate of growth in health care costs.

— Excerpt from speech by Romney on 5/25/12 in Des Moines, IA
Towards a measurement model

1. **Stage 1: Build a “dictionary” of cue lexicons**
   - Define the ideological structure.
   - Identify prototypical examples of each ideology → Cue terms that are indicative of ideologies.
Towards a measurement model

1. Stage 1: Build a “dictionary” of cue lexicons
   - Define the ideological structure.
   - Identify prototypical examples of each ideology → Cue terms that are indicative of ideologies.

2. Stage 2: Analyze speeches using learned ideological cue lexicons.
   - Build probabilistic sequence model to infer latent ideological states for each cue term in a speech.
   - Model should match our assumptions about cue-fillers and make use of cue lexicon.
Stage 1: Building “dictionary” of cue lexicons

Ideological books corpus

- The New Republic: Myth of American Decline
- We Can All Do Better
- God’s Politics
- Pitchforks and Torches
- The Conscience of a Liberal
- Noam Chomsky: Power in Systems
- Occupy the Economy: Challenging Capitalism

Cue lexicons:
- Center Left
- Religious Left
- Progressive
- Left
- Far Left
- Center
- Right
- Religious Right
- Libertarian
- Populist
- Far Right
Stage 1: Building “dictionary” of cue lexicons

Ideological books corpus
Cue discovery model

- We would like
  1. to take into account for the hierarchical nature of ideological structure,
  2. a sparse set of cue terms.

- Solution: Sparse Additive GenerativE model for text (Eisenstein et al., 2011) — SAGE
Stage 1: Building “dictionary” of cue lexicons

Cue discovery model

Sparse additive effects for cue discovery (Example)

Bag of terms probability

\[ p(w \mid \text{metadata}) \propto \prod_{w \in d} \exp \left( \eta_{\text{background}}w + \eta_{\text{left}}w + \eta_{\text{far left}}w + \eta_{\text{edu}}w + \eta_{\text{power systems}}w \right) \]
## Stage 1: Building “dictionary” of cue lexicons

### Cue discovery model

<table>
<thead>
<tr>
<th>Group</th>
<th>Authors</th>
<th>Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Center-Right</strong></td>
<td>D. Frum, M. McCain, C. T. Whitman (1,450)</td>
<td>governor bush; class voter; health care; republican president; george bush; state police; move forward; miss america; middle eastern; water buffalo; fellow citizens; sam’s club; american life; working class; general election; culture war; status quo; human dignity; same-sex marriage</td>
</tr>
<tr>
<td><strong>Libertarian</strong></td>
<td>Rand Paul, John Stossel, <em>Reason</em> (2,268)</td>
<td>medical marijuana; raw milk; rand paul; economic freedom; health care; government intervention; market economies; commerce clause; military spending; government agency; due process; drug war; minimum wage; federal law; ron paul; private property</td>
</tr>
<tr>
<td><strong>Religious Right</strong></td>
<td>(960)</td>
<td>daily saint; holy spirit; matthew [c/v]; john [c/v]; jim wallis; modern liberals; individual liberty; god’s word; jesus christ; elementary school; natural law; limited government; emerging church; private property; planned parenthood; christian nation; christian faith</td>
</tr>
</tbody>
</table>

Browse results at [http://www.ark.cs.cmu.edu/CLIP/](http://www.ark.cs.cmu.edu/CLIP/) or [http://is.gd/etchasketch](http://is.gd/etchasketch)
Each speech is generated by a hidden Markov model:

- ideologies — hidden states,
- cue terms — emitted,
HMM with a twist

Right

repeal

ObamaCare

Progressive (L)

social

security
HMM with a twist

Walk in ideology tree:

\[ p_{tree}(s_j \mid s_i; \zeta, \theta) = \left( \prod_{\langle u, v \rangle \in Path(s_i, s_j)} (1 - \zeta_u) \theta_u, v \right) \zeta_{s_j} \]
Stage 2: Cue Lag Ideological Proportions

HMM with a twist

Right

repeal

Obamacare

Progressive (L)

social

security

lag=7

Restart switch — noisy-OR model:

\[ p_{restart}(\text{continue} \mid \text{lag}) = (1 - \rho)^{\text{lag}+1} \]
An informed emission

\[ \text{Progressive (L)} \]

\[ \text{social security} \]

Informed Dirichlet prior for emission distribution

\[ \text{social security} \mid \text{Progressive} \sim \text{Dirichlet}(\beta^{\text{Progressive}}) \]

where

\[ \beta^{\text{Progressive}}_{\text{social security}} = \begin{cases} 
\beta_{\text{cue}} & \text{if social security } \in \text{Progressive cue lexicon} \\
\beta_{\text{default}} & \text{if social security } \in \text{any other cue lexicon}
\end{cases} \]
Learning and Inference

• Gibbs sampling to pick latent (ideology, restart?) variables
• Slice sampling for hyperparameters
Learning and Inference

- Gibbs sampling to pick latent (ideology, restart?) variables
- Slice sampling for hyperparameters

Ideological proportions

$$\sum_{\text{cue term}} p(\text{ideology} | \text{cue term}) \times \frac{1}{2} \left( \frac{\text{Lag before and after cue term}}{\text{Total lag}} \right)$$
Presidential campaign speeches

- Campaign speeches of 21 candidates each separated into epochs: primary and general elections

- Run model on each candidate separately with
  - independent transition parameters for each epoch, but
  - shared emission parameters for a candidate.
Mitt Romney
Mitt Romney
Barack Obama

Measuring Ideological Proportions in Political Speeches (EMNLP 2013) 21 / 28
Barack Obama

Measuring Ideological Proportions in Political Speeches (EMNLP 2013)
John McCain

Measuring Ideological Proportions in Political Speeches (EMNLP 2013) 22 / 28
John McCain
Evaluation?

- Is there a “gold standard”?
Evaluation?

- Is there a “gold standard”?

**Pre-registered hypothesis**

Statements made wrt domain experts’ *expectations* of the model’s output.
## Preregistered hypothesis

### Hypotheses

<table>
<thead>
<tr>
<th>Sanity checks (strong):</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S1. Republican primary candidates should tend to draw more from <strong>RIGHT</strong> than from <strong>LEFT</strong>.</td>
<td></td>
</tr>
<tr>
<td>S2. Democratic primary candidates should tend to draw more from <strong>LEFT</strong> than from <strong>RIGHT</strong>.</td>
<td></td>
</tr>
<tr>
<td>S3. In general elections, Democrats should draw more from the <strong>LEFT</strong> than the Republicans and vice versa for the <strong>RIGHT</strong>.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary hypotheses (strong):</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1. Romney, McCain and other Republicans should almost never draw from <strong>FAR LEFT</strong>, and extremely rarely from <strong>PROGRESSIVE</strong>.</td>
<td></td>
</tr>
<tr>
<td>P2. Romney should draw more heavily from the <strong>RIGHT</strong> than Obama in both stages of the 2012 campaign.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary hypotheses (moderate):</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P3. Romney should draw more heavily on words from the <strong>LIBERTARIAN</strong>, <strong>POPULIST</strong>, <strong>RELIGIOUS RIGHT</strong>, and <strong>FAR RIGHT</strong> in the primary compared to the general election. In the general election, Romney should draw more heavily on <strong>CENTER</strong>, <strong>CENTER-RIGHT</strong> and <strong>LEFT</strong> vocabularies.</td>
<td></td>
</tr>
</tbody>
</table>
Baselines

- Compare against “simplified” versions of the model:
  - **HMM**: Traditional HMM without ideological tree structure
  - **NoRes**: Weaker Markovian assumptions — never restart
  - **Mix**: Stronger Markovian assumptions — always restart
## Results

<table>
<thead>
<tr>
<th></th>
<th>CLIP</th>
<th>HMM</th>
<th>MIX</th>
<th>NoRes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanity checks</td>
<td>20/21</td>
<td>19/22</td>
<td>21/22</td>
<td>17/22</td>
</tr>
<tr>
<td>Strong hypotheses</td>
<td>31/34</td>
<td>23/33</td>
<td>28/34</td>
<td>30/34</td>
</tr>
<tr>
<td>Moderate hypotheses</td>
<td>14/17</td>
<td>14/17</td>
<td>12/17</td>
<td>11/17</td>
</tr>
<tr>
<td>Total</td>
<td>65/72</td>
<td>56/72</td>
<td>61/73</td>
<td>58/73</td>
</tr>
</tbody>
</table>
Conclusions

- Learn cue terms from lightly labeled books corpus.
  - Incorporating the structure of the ideology “tree”.

- Cue Lag Ideological Proportions model (CLIP)
  - Empirical support for the “etch-a-sketch” hypothesis.

- Evaluation framework with pre-registered hypotheses.

More model results at http://www.ark.cs.cmu.edu/CLIP/ or http://is.gd/etchasketch
Thank you

Output of model can be found at http://www.ark.cs.cmu.edu/CLIP/
or http://is.gd/etchasketch
Stage 1: Building “dictionary” of cue lexicons

SAGE for cue discovery

Sparse additive effects for cue discovery (Example)

Well, that gets technical, but there’s very exciting work going on

Bag of terms probability

\[
p(w \mid A(d); \eta) \propto \prod_{w \in d} \exp \left( \eta_{\text{background}} w + \eta_{\text{left}} w + \eta_{\text{far left}} w + \eta_{\text{edu}} w + \eta_{\text{power systems}} w \right)
\]

are worth studying, but universal grammar is the study of the genetic basis for language, the genetic basis of the language faculty. There can’t be any serious doubt that something like that exists. Otherwise an infant couldn’t reflexively acquire language from whatever complex data is around. So that’s not controversial. The only question is what the genetic basis of the language faculty is.

Sparse additive effects for cue discovery (Example)

Well, that gets technical, but there’s very exciting work going on

Bag of terms probability

\[ p(w \mid A(d); \eta) \propto \prod_{w \in d} \exp \left( \eta_{\text{background}} + \eta_{\text{left}} + \eta_{\text{far left}} + \eta_{\text{edu}} + \eta_{\text{power systems}} \right) \]

...are worth studying, but universal grammar is the study of the genetic basis for language, the genetic basis of the language faculty. There can’t be any serious doubt that something like that exists. Otherwise an infant couldn’t reflexively acquire language from whatever complex data is around. So that’s not controversial. The only question is what the genetic basis of the language faculty is.

Well, that gets technical, but there’s very exciting work going on refining the proposed principles of universal grammar. The concept is widely misunderstood in the media and in public discussions. Universal grammar is something different: it is not a set of universal observations about language. In fact, there are interesting generalizations about language that are worth studying, but universal grammar is the study of the genetic basis for language, the genetic basis of the language faculty. There can’t be any serious doubt that something like that exists. Otherwise an infant couldn’t reflexively acquire language from whatever complex data is around. So that’s not controversial. The only question is what the genetic basis of the language faculty is.

Sparse additive effects for cue discovery (Example)

Well, that gets technical, but there’s very exciting work going on.

Bag of terms probability

\[ p(w | A(d); \eta) \propto \prod_{w \in d} \exp \left( \eta_{\text{background}}^{w} + \eta_{\text{left}}^{w} + \eta_{\text{far left}}^{w} + \eta_{\text{edu}}^{w} + \eta_{\text{power systems}}^{w} \right) \]

are worth studying, but universal grammar is the study of the genetic basis for language, the genetic basis of the language faculty. There can’t be any serious doubt that something like that exists. Otherwise an infant couldn’t reflexively acquire language from whatever complex data is around. So that’s not controversial. The only question is what the genetic basis of the language faculty is.

Sparse additive effects for cue discovery (Example)

Well, that gets technical, but there’s very exciting work going on

Bag of terms probability

\[ p(w \mid A(d); \eta) \propto \prod_{w \in d} \exp \left( \eta_{\text{background}} + \eta_{\text{left}} + \eta_{\text{far left}} + \eta_{\text{edu}} + \eta_{\text{power systems}} \right) \]

are worth studying, but universal grammar is the study of the genetic basis for language, the genetic basis of the language faculty. There can’t be any serious doubt that something like that exists. Otherwise an infant couldn’t reflexively acquire language from whatever complex data is around. So that’s not controversial. The only question is what the genetic basis of the language faculty is.

Sparse additive effects for cue discovery (Example)

Well, that gets technical, but there’s very exciting work going on

**Bag of terms probability**

\[
p(w \mid A(d); \eta) \propto \prod_{w \in d} \exp \left( \eta_{w}^{\text{background}} + \eta_{w}^{\text{left}} + \eta_{w}^{\text{far left}} + \eta_{w}^{\text{edu}} + \eta_{w}^{\text{power systems}} \right)
\]

**SAGE objective**

\[
\max_{\eta} \sum_{d} \sum_{w \in d} \log p(w \mid A(d); \eta) - \sum_{a \in A} \lambda_{a} \| \eta^{a} \|_{1}
\]

the genetic basis of the language faculty is.

Sparse additive effects for cue discovery

Word distribution

\[
p(w \mid A(d); \eta) = \frac{\exp \left( \eta^0_w + \sum_{a \in A(d)} \eta^a_w \right)}{Z(A(d), \eta)}
\]

\(a \in A(d)\) — attributes of document \(d\).
\(\eta^0\) — background (log) frequencies of terms in the corpus.

Possible attributes/effects are:
\(\eta^\text{left}, \eta^\text{center}, \eta^\text{right}\) — coarse ideology effect.
\(\eta^\text{prgrsv}, \eta^\text{lbtn}, \text{etc}\) — fine ideology effect, corresponding to the fine grained ideologies in our tree
\(\eta^\text{topic}\) — topical effect. Book chapters in our corpus are manually assigned 1 of 61 topics.
\(\eta^d\) — document specific effect, which captures idiosyncratic usage within a single document.
Sparse additive effects for cue discovery

**SAGE objective**

$$\max_{\eta} \sum_d \sum_{w \in d} \log p(w \mid \mathcal{A}(d); \eta) - \sum_{a \in \mathcal{A}} \lambda_a \| \eta^a \|_1$$

$\mathcal{A}(d)$ — set of attributes/effects in document $d$.

$\lambda_a$ — regularization for attribute $a$, controls the sparsity of $\eta^a$. 
Validation

bigrams, trigrams, 4-grams

unigrams

PMI  SAGE  WAPMI  SAGE w/ topics