Measuring Latent Political Ideal Points of Twitter Users from User Description Text Data

In Son, Patrick, Yanwen

University of Michigan

insonz@umich.edu, pywu@umich.edu, yanwenzh@umich.edu

December 11, 2018

Agenda

Background

- **Definition of Political Ideal Point**: A point which places an individual along a latent left-right political continuum that explains or predicts his/her political behavior

Background

- **Definition of Political Ideal Point**: A point which places an individual along a latent left-right political continuum that explains or predicts his/her political behavior
- **Example**: Bernie Sanders is more left than Dianne Feinstein because he takes on more liberal positions and votes more liberally, even though both individuals are in the Democratic Party
- **Prior research**: Barbera (2015), Simon, Jackman, Rivers (2004), Poole (2005), Bonica (2014)

Background

- Most of the literature has focused on estimating ideal points for politicians.

Background

- Most of the literature has focused on estimating ideal points for politicians.
- **Project Goal**: Estimate latent political ideal points of ordinary Twitter users using Bayesian estimation techniques
- We will focus on estimating this latent political ideal point using text from the Twitter user biographies

Methodology: Intuitive Approach

- Specifically, we focus on extracting specific political keywords that users may put in their biographies as indicators of political affiliations
- **Example**: Let's say that we were interested in the words "Clinton" and "Trump." If user *i*'s biography is, "I love Donald Trump!", then $y_{i,trump} = 1$ and $y_{i,clinton} = 0$.

Methodology: Technical Details

- The Bayesian model we develop largely resembles a combination of models found in Barbera (2015), Simon, Jackman, and Rivers (2004), and Hoff, Raftery, and Handcock (2002)
- Its closest analogue is a Bayesian item-response theory model

Methodology: Technical Details

- Suppose that each Twitter user is presented with a choice to mention or not mention a political keyword, which is a word that clearly demarcates a political stance or affiliation.
- Let $y_{ij} = 1$ if user i mentions word j in their biography, and let $y_{ij} = 0$ otherwise.

Methodology: Technical Details

- Suppose that each Twitter user is presented with a choice to mention or not mention a political keyword, which is a word that clearly demarcates a political stance or affiliation.
- Let $y_{ij} = 1$ if user i mentions word j in their biography, and let $y_{ij} = 0$ otherwise.
- We can consider this the function of the squared Euclidean distance in the latent political dimension between user i and word j: $-\gamma(\theta_i - \phi_j)^2$, where $\theta_i \in \mathbb{R}$ is the latent political ideal point of Twitter user i along this latent political dimension, ϕ_j is the political ideal point of word j along this political dimension, and γ is the discrimination parameter, or how important this relationship is to estimating the political ideal point.

Methodology: Technical Details

- Let β_i be a measure of how political an individual is on Twitter. Sometimes individuals may spam political words in their autobiographies, while others may only mention a single political keyword.

Methodology: Technical Details

- Let β_i be a measure of how political an individual is on Twitter. Sometimes individuals may spam political words in their autobiographies, while others may only mention a single political keyword.
- Then, assuming conditional independence between users, our likelihood in this model is

$$p(\mathbf{y}|\theta, \phi, \beta, \gamma) = \prod_{i=1}^{n} \prod_{j=1}^{m} (logit^{-1}(\pi_{ij}))^{y_{ij}} (1 - logit^{-1}(\pi_{ij}))^{1-y_{ij}}$$

where $\pi_{ij} = \beta_i - \gamma(\theta_i - \phi_j)^2$. Then, the full posterior is

$$p(\theta, \phi, \beta, \gamma|y) \propto p(\mathbf{y}|\theta, \phi, \beta, \gamma)p(\theta, \phi, \beta, \gamma)$$

Methodology: Technical Details

- We assume the following priors: $\beta_i \sim N(\mu_\beta, \sigma^2_\beta)$, $\theta_i \sim N(\mu_\theta, \sigma^2_\theta)$, and $\phi_j \sim N(\mu_\phi, \sigma^2_\phi)$.
- Then, the full joint posterior distribution is

$$p(\theta, \phi, \beta, \gamma | y) \propto \prod_{i=1}^{n} \prod_{j=1}^{m} (logit^{-1}(\pi_{ij}))^{y_{ij}} (1 - logit^{-1}(\pi_{ij}))^{1-y_{ij}} \times$$

$$\times \prod_{i=1}^{n} N(\beta_i | \mu_\beta, \sigma^2_\beta) \prod_{i=1}^{n} N(\theta_i | \mu_\theta, \sigma^2_\theta) \prod_{j=1}^{m} N(\phi_j | \mu_\phi, \sigma^2_\phi)$$

Methodology: Now in Plain English

- We want to estimate θ_i for each user i. This is each user's ideal point along a latent left-right political continuum.
- Our assumption is that the closer the ideal point of user i to the ideal point of word j *along the same latent left-right political continuum*, the more likely user i will use word j in his or her autobiography.
- Everything else is just technical details.

Why a Bayesian Approach?

- The number of parameters is very large (one β for each user, one θ for each user, one ϕ for each word), so a Bayesian approach turns what is typically a very difficult problem in classical estimation to a routine application of MCMC.
- It also allows us to incorporate previous knowledge through other studies of the distribution of ideal points of ordinary citizens through the priors. See Barbera (2015) and Bonica (2014).

Data

- Our Twitter data comes from dissertation work of Patrick Wu.
- It was collected in the month before the November 8, 2016 general U.S. election.
- All users in this dataset use at least one of the 14 political keywords we selected, as detailed in the next slide.
- There are $9,190$ user biographies in our dataset.
- To get matches, we stemmed all words in the user biographies and matched based on stemmed words.
- Thus, we are estimating 18,396 parameters.

Data: Keyword Selection

We are analyzing 14 keywords from Twitter autobiographies:

Trump	Republican	MAGA	AlwaysTrump
Clinton	Democrat	StrongerTogether	ImWithHer
Donald	RealDonaldTrump	NeverTrump	
Hillary	HillaryClinton	NeverHillary	

Posterior Distribution for Keywords (ϕ_j)

The following table tells the posterior distribution for each keywords ϕ_j (mean value, 0.01, 0.5, 0.99 quantiles and 95% credible interval):

Keyword	mean	0.01	**95% cred. interval**	median	0.99	SD
Trump	2.199	2.068	(2.085,2.318)	2.198	2.340	0.060
Clinton	-4.594	-4.832	(-4.791,-4.415)	-4.592	-4.384	0.096
Donald	4.686	4.477	(4.508,4.875)	4.683	4.914	0.096
Hillary	-3.757	-3.947	(-3.920,-3.600)	-3.756	-3.578	0.082
Republican	4.017	3.836	(3.859,4.186)	4.016	4.221	0.084
Democrat	-3.097	-3.274	(-3.246,-2.957)	-3.096	-2.933	0.073
RealDonaldTrump	5.411	5.167	(5.207, 5.628)	5.410	5.675	0.110
HillaryClinton	-4.874	-5.114	(-5.074,-4.683)	-4.872	-4.652	0.100
MAGA	3.831	3.651	(3.679,3.993)	3.829	4.021	0.080
StrongerTogether	-5.200	-5.454	(-5.414,-4.993)	-5.197	-4.961	0.108
NeverHillary	4.389	4.189	(4.216, 4.574)	4.387	4.604	0.090
NeverTrump	-4.145	-4.356	(-4.323,-3.981)	-4.143	-3.951	0.088
AlwaysTrump	6.476	6.141	(6.189,6.789)	6.473	6.857	0.151
ImWithHer	-3.166	-3.338	(-3.311,-3.030)	-3.164	-3.005	0.072

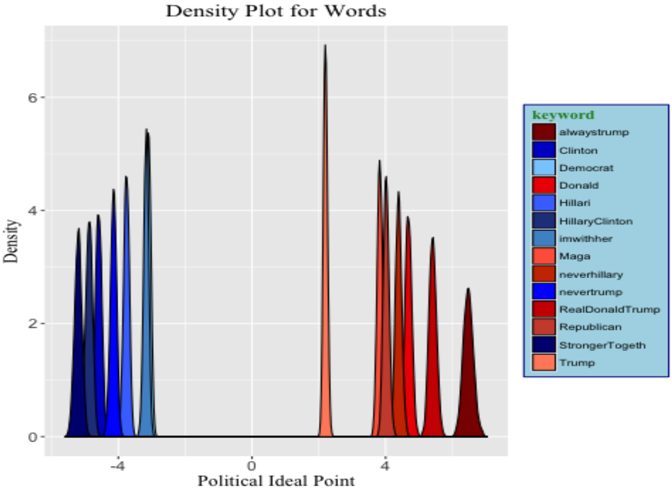
Posterior Distribution for Keywords (ϕ_j)

The following table tells the posterior distribution for each keywords ϕ_j (mean value, 0.01, 0.5, 0.99 quantiles and 95% credible interval):

Keyword	mean	0.01	**95% cred. interval**	median	0.99	SD
Trump	2.199	2.068	(2.085,2.318)	2.198	2.340	0.060
Clinton	-4.594	-4.832	(-4.791,-4.415)	-4.592	-4.384	0.096
Donald	4.686	4.477	(4.508,4.875)	4.683	4.914	0.096
Hillary	-3.757	-3.947	(-3.920,-3.600)	-3.756	-3.578	0.082
Republican	4.017	3.836	(3.859,4.186)	4.016	4.221	0.084
Democrat	-3.097	-3.274	(-3.246,-2.957)	-3.096	-2.933	0.073
RealDonaldTrump	5.411	5.167	(5.207, 5.628)	5.410	5.675	0.110
HillaryClinton	-4.874	-5.114	(-5.074,-4.683)	-4.872	-4.652	0.100
MAGA	3.831	3.651	(3.679,3.993)	3.829	4.021	0.080
StrongerTogether	-5.200	-5.454	(-5.414,-4.993)	-5.197	-4.961	0.108
NeverHillary	4.389	4.189	(4.216, 4.574)	4.387	4.604	0.090
NeverTrump	-4.145	-4.356	(-4.323,-3.981)	-4.143	-3.951	0.088
AlwaysTrump	6.476	6.141	(6.189,6.789)	6.473	6.857	0.151
ImWithHer	-3.166	-3.338	(-3.311,-3.030)	-3.164	-3.005	0.072

- **AlwaysTrump** goes to the most positive side, whereas **StrongerTogether** goes to the most negative side.

Posterior Distribution for Keywords(ϕ_j)



1. **Left:** Democratic and **Right:** Republican party affiliated words.

Posterior Distributions for Parameters (β_i, θ_i and γ)

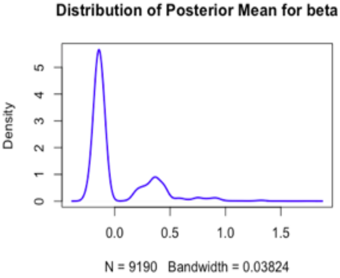
The following table tells the posterior distribution of individual effects (β_i, θ_i) and discrimation parameter (γ):

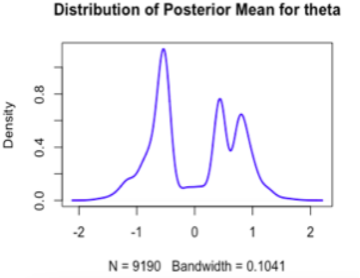
Parameter	mean	0.01	**95% cred. interval**	median	0.99	SD
β_{avg}	0.01	-1.80	(-1.50,1.40)	0.03	1.64	0.74
θ_{avg}	0.00	-1.64	(-1.37,1.36)	0.00	1.63	0.70
γ	0.18	0.17	(0.17,0.20)	0.18	0.20	0.01

Posterior Distributions for Parameters (β_i, θ_i and γ)

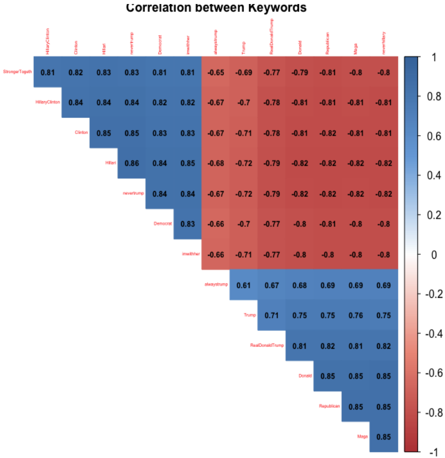
The following table tells the posterior distribution of individual effects (β_i, θ_i) and discrimation parameter (γ):

Parameter	mean	0.01	**95% cred. interval**	median	0.99	SD
β_{avg}	0.01	-1.80	(-1.50,1.40)	0.03	1.64	0.74
θ_{avg}	0.00	-1.64	(-1.37,1.36)	0.00	1.63	0.70
γ	0.18	0.17	(0.17,0.20)	0.18	0.20	0.01





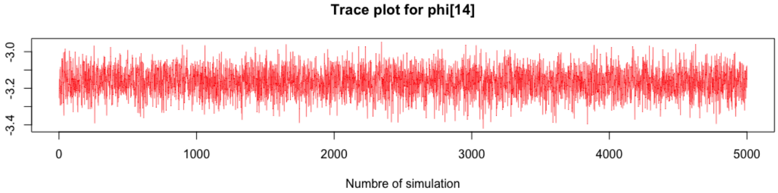
The Relationship Between Keywords



1. The correlation of ϕ_j appears to be **positive/negative** if the two keywords are affiliated with the **same/opposing** party in prediction.

Model Diagnostics

1. Trace plot



2. Geweke diagnostic test

ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
-1.015	1.160	-1.105	0.971	-0.804	0.909	-1.054
ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{14}
1.101	-1.158	1.048	-1.148	0.929	-1.661	0.929

Model Diagnostics

- Geweke diagnostic test statistics for 18396 parameters

1%	2.5%	25%	50%	75%	97.5%	99%
-2.318	-1.950	-0.687	-0.014	0.664	1.936	2.350

- pD: We obtained pD: 15830.61 with 18396 parameters in our model, so $\frac{pD}{\#\ parameter} < 1$.
- Gelman-Rubin statistic: currently having trouble running on Flux...

Validation: Predicting Retweets of Dem. and Rep. Accounts

- Although the confidence intervals on the θ_i values are quite large, we think their width comes from our small n.
- We have the number of retweets from popular Democratic/left-leaning accounts and the total number of retweets from popular Republican/right-leaning accounts for each individual in our dataset.

Validation: Predicting Retweets of Dem. and Rep. Accounts

- Although the confidence intervals on the θ_i values are quite large, we think their width comes from our small n.
- We have the number of retweets from popular Democratic/left-leaning accounts and the total number of retweets from popular Republican/right-leaning accounts for each individual in our dataset.
- Because many individuals do not have retweets, and because of overdispersion concerns, we use a zero-inflated negative binomial model.
- Here, the dependent variable is the number of left or right retweets, and the independent variable is the θ values.

Validation: Predicting Retweets of Dem. and Rep. Accounts

Count	*Dem. Accts*	*Rep. Accts*
(Intercept)	3.12 (0.02)	4.49 (0.02)
θ	-1.22 (0.03)	1.57 (0.03)
log(theta)	-0.65 (0.02)	-1.00 (0.02)
Zero-Inflated		
(Intercept)	4.48 (0.17)	5.18 (0.20)
θ	2.25 (0.08)	-2.03 (0.12)
log(1 + RT Count)	-0.91 (0.03)	-1.27 (0.04)
log likelihood	-27870	-38090

- log(theta) denotes the overdispersion parameter.

Validation: Predicting Other Types of Twitter Behavior

- We find that this pattern holds for retweets of Democratic members of Congress and retweets of Republican members of Congress
- We find that this pattern holds for retweets of Clinton vs. Trump
- We find that this pattern holds for the usage of hashtags typically associated with Democrats and for the usage of hashtags typically associated with Republicans
- Lastly, we find that this pattern also holds for favorites of popular Democratic accounts and favorites of popular Republican accounts.

Conclusion and Future Work

- This method does a good job placing words on the expected side of the political continuum.

Conclusion and Future Work

- This method does a good job placing words on the expected side of the political continuum.
- The ideal points estimated for each individual through θ_i are predicting other political behaviors on Twitter.

Conclusion and Future Work

- This method does a good job placing words on the expected side of the political continuum.
- The ideal points estimated for each individual through θ_i are predicting other political behaviors on Twitter.
- Extend model to people who may not use political keywords.

Conclusion and Future Work

- This method does a good job placing words on the expected side of the political continuum.
- The ideal points estimated for each individual through θ_i are predicting other political behaviors on Twitter.
- Extend model to people who may not use political keywords.
- Also, in the future we can implement a Metropolis-Hasting approach that fixes the word ideal points and updates ideal points for individuals only for faster computational times.


Any
Questions

JAGS Implementation

```
JAGS_ideal_points <- function(){
  #Prior
  for(j in 1:J){
    beta[j] ~ dnorm(0,1)
  }
  for(k in 1:K){
    phi[k] ~ dnorm(phi_mu,phi_tau)
  }
  for(j in 1:J){
    theta[j] ~ dnorm(0,1)
  }
  phi_mu ~ dunif(-99999999,99999999)
  phi_sigma ~ dunif(-99999999,99999999)
  phi_tau <- pow(phi_sigma,-2)

  gamma ~ dunif(-99999999,99999999)

  #Likelihood
  for (j in 1:J){
    for(k in 1:K){
      Y[j,k] ~ dbern(prob[j,k])
      logit(prob[j,k]) <- beta[j] - gamma*pow(theta[j] - phi[k],2)
    }
  }
}
```