Text as Data

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

November 20th, 2014

Ideological Scaling

1) Task
   - Measure political actors' position in policy space
   - Low dimensional representation of beliefs
2) Objective function
   - Linear Discriminant Analysis (ish)$\rightsquigarrow$ Wordscores (today)
   - Item Response Theory $\rightsquigarrow$ Wordfish
   - Item Response Theory + Roll Call Votes $\rightsquigarrow$ Issue-specific ideal points (12/2)
3) Optimization
   - Wordscores$\rightsquigarrow$ straightforward, based on training texts
   - Wordfish$\rightsquigarrow$ EM, MCMC methods
4) Validation
   - What is the goal of embedding?
   - What is the gold standard?

The Spatial Model

- Suppose we have actor i $(i = 1, 2, 3, \ldots, N)$

The Spatial Model

- Suppose we have actor i ($i = 1, 2, 3, \ldots, N$)
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$

The Spatial Model

- Suppose we have actor i $(i = 1, 2, 3, \ldots, N)$
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$
- We describe actor i's utility from proposal $\boldsymbol{p} \in \Re^M$ with utility function

The Spatial Model

- Suppose we have actor i ($i = 1, 2, 3, \ldots, N$)
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$
- We describe actor i's utility from proposal $\boldsymbol{p} \in \Re^M$ with utility function

$$u_i(\boldsymbol{\theta}_i, \boldsymbol{p}) = -d(\boldsymbol{\theta}_i, \boldsymbol{p})$$

The Spatial Model

- Suppose we have actor i ($i = 1, 2, 3, \ldots, N$)
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$
- We describe actor i's utility from proposal $\boldsymbol{p} \in \Re^M$ with utility function

$$
\begin{aligned}
u_i(\boldsymbol{\theta}_i, \boldsymbol{p}) &= -d(\boldsymbol{\theta}_i, \boldsymbol{p}) \\
&= -\sum_{l=1}^{L} (\underbrace{\theta_{il}}_{\text{ideal policy}} - p_l)^2
\end{aligned}
$$

The Spatial Model

- Suppose we have actor i ($i = 1, 2, 3, \ldots, N$)
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$
- We describe actor i's utility from proposal $\boldsymbol{p} \in \Re^M$ with utility function

$$\begin{aligned} u_i(\boldsymbol{\theta}_i, \boldsymbol{p}) &= -d(\boldsymbol{\theta}_i, \boldsymbol{p}) \\ &= -\sum_{l=1}^{L} (\underbrace{\theta_{il}}_{\text{ideal policy}} - p_l)^2 \end{aligned}$$

Estimation goal: $\widehat{\boldsymbol{\theta}}_i$

The Spatial Model

- Suppose we have actor i ($i = 1, 2, 3, \ldots, N$)
- Actor has ideal point $\boldsymbol{\theta}_i \in \Re^M$
- We describe actor i's utility from proposal $\boldsymbol{p} \in \Re^M$ with utility function

$$
\begin{aligned}
u_i(\boldsymbol{\theta}_i, \boldsymbol{p}) &= -d(\boldsymbol{\theta}_i, \boldsymbol{p}) \\
&= -\sum_{l=1}^{L} (\underbrace{\theta_{il}}_{\text{ideal policy}} - p_l)^2
\end{aligned}
$$

Estimation goal: $\widehat{\boldsymbol{\theta}}_i$

Scaling$\rightsquigarrow$ placing actors in low-dimensional space (like principal components!)

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
    - Roll Call Data $\rightsquigarrow$ 1789-Present

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
  - Roll Call Data $\rightsquigarrow$ 1789-Present
  - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
  - Roll Call Data $\rightsquigarrow$ 1789-Present
  - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology
- Wildly successful:

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
  - Roll Call Data $\rightsquigarrow$ 1789-Present
  - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology
- Wildly successful:
  - Estimates are accurate: face validity Congressional scholars agree upon

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
  - Roll Call Data $\rightsquigarrow$ 1789-Present
  - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology
- Wildly successful:
  - Estimates are accurate: face validity Congressional scholars agree upon
  - Insightful $\rightsquigarrow$ unidimensional Congress

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
    - Roll Call Data $\rightsquigarrow$ 1789-Present
    - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology
- Wildly successful:
    - Estimates are accurate: face validity Congressional scholars agree upon
    - Insightful $\rightsquigarrow$ unidimensional Congress
    - Extensible: insight of IRT allows model to be embedded in many forms

Estimating Ideal Points: Roll Call Data and the US Congress

US Congress and Roll Call

- Poole and Rosenthal `voteview`
  - Roll Call Data $\rightsquigarrow$ 1789-Present
  - `NOMINATE` methods $\rightsquigarrow$ place legislators on one dimension, estimate of ideology
- Wildly successful:
  - Estimates are accurate: face validity Congressional scholars agree upon
  - Insightful $\rightsquigarrow$ unidimensional Congress
  - Extensible: insight of IRT allows model to be embedded in many forms
  - Widely used: hard to write a paper on American political institutions with ideal points

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
   - Weak party pressure $\rightsquigarrow$ individual discretion on votes

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
   - Weak party pressure $\rightsquigarrow$ individual discretion on votes
   - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
   - Weak party pressure $\rightsquigarrow$ individual discretion on votes
   - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.
   - Spirling and Quinn (2011) $\rightsquigarrow$ mixture model like models for blocs in UK Parliament

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
    - Weak party pressure $\rightsquigarrow$ individual discretion on votes
    - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.
    - Spirling and Quinn (2011) $\rightsquigarrow$ mixture model like models for blocs in UK Parliament
2) Not everyone votes!

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
   - Weak party pressure $\rightsquigarrow$ individual discretion on votes
   - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.
   - Spirling and Quinn (2011) $\rightsquigarrow$ mixture model like models for blocs in UK Parliament
2) Not everyone votes!
   - Voters $\rightsquigarrow$ survey responses (but problems with that)

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
    - Weak party pressure $\rightsquigarrow$ individual discretion on votes
    - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.
    - Spirling and Quinn (2011) $\rightsquigarrow$ mixture model like models for blocs in UK Parliament
2) Not everyone votes!
    - Voters $\rightsquigarrow$ survey responses (but problems with that)
    - Challengers $\rightsquigarrow$ NPAT surveys (but they don't fill those out anymore)

Estimating Ideal Points in General

Two Limitations with the NOMINATE project:

1) US Congress is distinct $\rightsquigarrow$ roll call votes fail to measure ideology in other settings
   - Weak party pressure $\rightsquigarrow$ individual discretion on votes
   - Parliamentary systems $\rightsquigarrow$ no discretion, no variation.
   - Spirling and Quinn (2011) $\rightsquigarrow$ mixture model like models for blocs in UK Parliament
2) Not everyone votes!
   - Voters $\rightsquigarrow$ survey responses (but problems with that)
   - Challengers $\rightsquigarrow$ NPAT surveys (but they don't fill those out anymore)
   - Bonica (2013, 2014) $\rightsquigarrow$ estimate ideology from donations (but not everyone donates)

Estimating Ideal Points in General

But Everyone talks!

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature $\rightsquigarrow$ relies upon intuition from US Congress

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature$\rightsquigarrow$ relies upon intuition from US Congress
    - Hard not to find ideology

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature$\rightsquigarrow$ relies upon intuition from US Congress
    - Hard not to find ideology
    - Behavior that is primarily ideological

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature$\rightsquigarrow$ relies upon intuition from US Congress
  - Hard not to find ideology
  - Behavior that is primarily ideological
- Reality: scaling is much more difficult than roll call voting examples

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature$\rightsquigarrow$ relies upon intuition from US Congress
  - Hard not to find ideology
  - Behavior that is primarily ideological
- Reality: scaling is much more difficult than roll call voting examples
  - Hard to find ideology

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature $\rightsquigarrow$ relies upon intuition from US Congress
  - Hard not to find ideology
  - Behavior that is primarily ideological
- Reality: scaling is much more difficult than roll call voting examples
  - Hard to find ideology
  - Much of political speech reveals little about position on ideological spectrum $\rightsquigarrow$ advertising, regional

Estimating Ideal Points in General

But Everyone talks!

- If we could scale based on conversation, we can measure ideology anywhere
- Much of the literature$\rightsquigarrow$ relies upon intuition from US Congress
    - Hard not to find ideology
    - Behavior that is primarily ideological
- Reality: scaling is much more difficult than roll call voting examples
    - Hard to find ideology
    - Much of political speech reveals little about position on ideological spectrum$\rightsquigarrow$ advertising, regional

Healthy skepticism!

Our plan

Our plan

1) Wordscores $\rightsquigarrow$ "supervised" scaling

Our plan

1) Wordscores $\rightsquigarrow$ "supervised" scaling
2) Wordfish $\rightsquigarrow$ single dimension

Wordscores $\rightsquigarrow$ Big in Europe

Wordscores, Like the Hoff $\rightsquigarrow$ Big in Europe

Wordscores: Objective Function

For each legislator i, suppose we observe D_i documents.

Wordscores: Objective Function

For each legislator i, suppose we observe D_i documents.
Define:

Wordscores: Objective Function

For each legislator i, suppose we observe D_i documents.
Define:

$$\boldsymbol{x}_i = \sum_{l=1}^{D_i} \boldsymbol{x}_{il}$$

Wordscores: Objective Function

For each legislator i, suppose we observe D_i documents.
Define:

$$
\begin{aligned}
\boldsymbol{x}_i &= \sum_{l=1}^{D_i} \boldsymbol{x}_{il} \\
&= \sum_{l=1}^{D_i} (x_{il1}, x_{il2}, \ldots, x_{ilJ})
\end{aligned}
$$

Wordscores: Objective Function

For each legislator i, suppose we observe D_i documents.
Define:

$$
\begin{aligned}
\boldsymbol{x}_i &= \sum_{l=1}^{D_i} \boldsymbol{x}_{il} \\
&= \sum_{l=1}^{D_i} (x_{il1}, x_{il2}, \ldots, x_{ilJ})
\end{aligned}
$$

$\boldsymbol{x}_i \rightsquigarrow$ aggregation across documents.

Wordscores: Objective Functions

Choose two legislators as exemplars

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

$$P_{jL} \quad = \quad \text{Probability of word from Liberal}$$

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

$$\begin{aligned} P_{jL} &= \text{Probability of word from Liberal} \\ P_{jC} &= \text{Probability of word from Conservative} \end{aligned}$$

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

$$\begin{aligned} P_{jL} &= \text{Probability of word from Liberal} \\ P_{jC} &= \text{Probability of word from Conservative} \end{aligned}$$

Define the score for word j

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

$$
\begin{aligned}
P_{jL} &= \text{Probability of word from Liberal} \\
P_{jC} &= \text{Probability of word from Conservative}
\end{aligned}
$$

Define the score for word j

$$S_j = Y_C P_{jC} + Y_L P_{jL}$$

Wordscores: Objective Functions

Choose two legislators as exemplars

- Legislator $L \in \{1, 2, \ldots, N\}$ is liberal. $Y_L = -1$
- For example, might select Elizabeth Warren
- Legislator $C \in \{1, 2, \ldots, N\}$ is Conservative. $Y_C = 1$
- For example, might select Ted Cruz

For each word j we can define:

$$\begin{aligned} P_{jL} &= \text{Probability of word from Liberal} \\ P_{jC} &= \text{Probability of word from Conservative} \end{aligned}$$

Define the score for word j

$$\begin{aligned} S_j &= Y_C P_{jC} + Y_L P_{jL} \\ &= P_{jC} - P_{jL} \end{aligned}$$

Wordscores: Objective Functions

Scale other legislators:

Wordscores: Objective Functions

Scale other legislators:

$$N_i = \sum_{j=1}^{J} \boldsymbol{x}_i$$

Wordscores: Objective Functions

Scale other legislators:

$$N_i = \sum_{j=1}^{J} \boldsymbol{x}_i$$

$\hat{\theta}_i$ is

Wordscores: Objective Functions

Scale other legislators:

$$N_i \quad = \quad \sum_{j=1}^{J} \boldsymbol{x}_i$$

$\hat{\theta}_i$ is

$$\hat{\theta}_i \quad = \quad \sum_{j=1}^{J} \left(\frac{x_{ij}}{N_i} \right) S_j$$

Wordscores: Objective Functions

Scale other legislators:

$$N_i = \sum_{j=1}^{J} \boldsymbol{x}_i$$

$\hat{\theta}_i$ is

$$\hat{\theta}_i = \sum_{j=1}^{J} \left(\frac{x_{ij}}{N_i}\right) S_j$$

$$= \frac{\boldsymbol{x}_i^{'}}{N_i}\boldsymbol{S}$$

Wordscores: Optimization

$$
\begin{aligned}
N_L &= \sum_{m=1}^{J} x_{mL} \\
N_C &= \sum_{m=1}^{J} x_{mC}
\end{aligned}
$$

Wordscores: Optimization

$$
\begin{aligned}
N_L &= \sum_{m=1}^{J} x_{mL} \\
N_C &= \sum_{m=1}^{J} x_{mC}
\end{aligned}
$$

Estimate P_{jL}, P_{jC}, and S_j

Wordscores: Optimization

$$
\begin{aligned}
N_L &= \sum_{m=1}^{J} x_{mL} \\
N_C &= \sum_{m=1}^{J} x_{mC}
\end{aligned}
$$

Estimate P_{jL}, P_{jC}, and S_j

$$
P_{jL} = \frac{\frac{x_{jL}}{N_L}}{\frac{x_{jL}}{N_L} + \frac{x_{jC}}{N_C}}
$$

Wordscores: Optimization

$$
\begin{aligned}
N_L &= \sum_{m=1}^{J} x_{mL} \\
N_C &= \sum_{m=1}^{J} x_{mC}
\end{aligned}
$$

Estimate P_{jL}, P_{jC}, and S_j

$$
\begin{aligned}
P_{jL} &= \frac{\frac{x_{jL}}{N_L}}{\frac{x_{jL}}{N_L} + \frac{x_{jC}}{N_C}} \\
P_{jC} &= 1 - P_{jL} = \frac{\frac{x_{jC}}{N_C}}{\frac{x_{jL}}{N_L} + \frac{x_{jC}}{N_C}}
\end{aligned}
$$

Wordscores: Optimization

$$
\begin{aligned}
N_L &= \sum_{m=1}^{J} x_{mL} \\
N_C &= \sum_{m=1}^{J} x_{mC}
\end{aligned}
$$

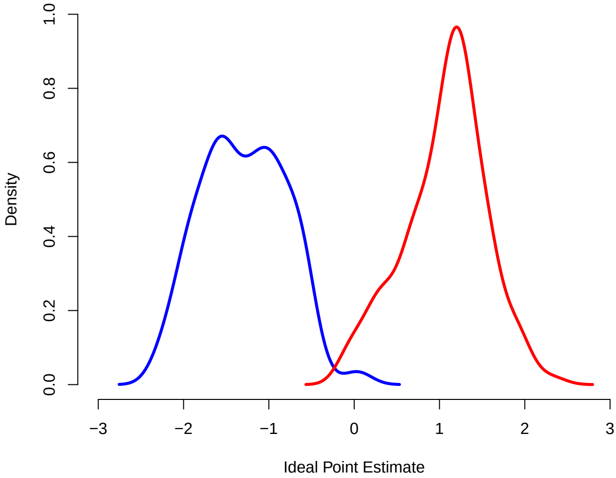
Estimate P_{jL}, P_{jC}, and S_j

$$
\begin{aligned}
P_{jL} &= \frac{\frac{x_{jL}}{N_L}}{\frac{x_{jL}}{N_L} + \frac{x_{jC}}{N_C}} \\
P_{jC} &= 1 - P_{jL} = \frac{\frac{x_{jC}}{N_C}}{\frac{x_{jL}}{N_L} + \frac{x_{jC}}{N_C}} \\
S_j &= P_{jC} - P_{jL}
\end{aligned}
$$

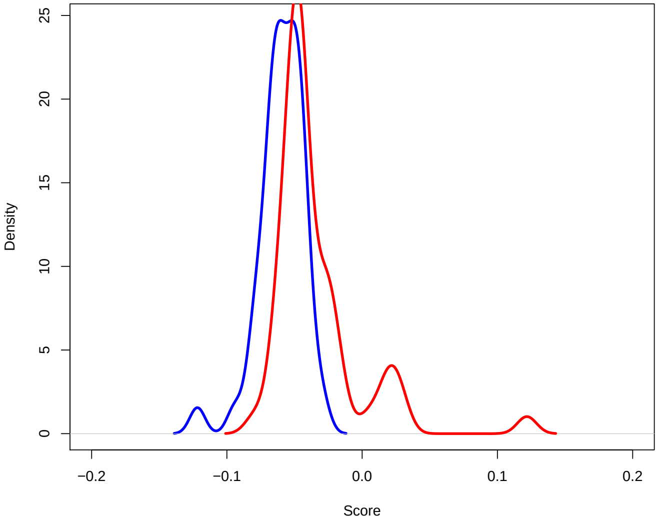
Applied to the Senate Press Releases

$L =$ Ted Kennedy
$C =$ Tom Coburn
Apply to other senators.

Applying to Senate Press Releases $\rightsquigarrow$ Gold Standard Scaling from NOMINATE



Applying to Senate Press Releases⇝ WordScores



Supervised Scaling

Wordscores is one example of supervised embedding:

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers
2) Multinomial Inverse Regression

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers
2) Multinomial Inverse Regression
3) Any method that separates words

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers
2) Multinomial Inverse Regression
3) Any method that separates words

The findings will depend heavily on supervision

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers
2) Multinomial Inverse Regression
3) Any method that separates words

The findings will depend heavily on supervision

- Sensitive to who is chosen

Supervised Scaling

Wordscores is one example of supervised embedding:

1) Linear Discriminant Analysis (LDA) $\rightsquigarrow$ the Federalist Papers
2) Multinomial Inverse Regression
3) Any method that separates words

The findings will depend heavily on supervision

- Sensitive to who is chosen
- False prediction problem $\rightsquigarrow$ speech accomplishes many goals, only some of which are ideological

What Can Go Wrong with Wordscores?

Lowe (2008): Discusses potentially problematic wordscores properties

What Can Go Wrong with Wordscores?

Lowe (2008): Discusses potentially problematic wordscores properties

1) Each word is weighted equally (fixable with different scoring procedure)

What Can Go Wrong with Wordscores?

Lowe (2008): Discusses potentially problematic wordscores properties

1) Each word is weighted equally (fixable with different scoring procedure)
2) Unique words are conflated with centrist (fixable with MCQ fightin' words style algorithm)

What Can Go Wrong with Wordscores?

Lowe (2008): Discusses potentially problematic wordscores properties

1) Each word is weighted equally (fixable with different scoring procedure)
2) Unique words are conflated with centrist (fixable with MCQ fightin' words style algorithm)
3) General problem: hard to interpret $\rightsquigarrow$ lack of a model

What Can Go Wrong with Wordscores?

Lowe (2008): Discusses potentially problematic wordscores properties

1) Each word is weighted equally (fixable with different scoring procedure)
2) Unique words are conflated with centrist (fixable with MCQ fightin' words style algorithm)
3) General problem: hard to interpret $\rightsquigarrow$ lack of a model

To be fair: fast, nonparametric, and novel [trailblazing] method for scoring documents (starts conversation)

Unsupervised Embedding

Basic idea:

Unsupervised Embedding

Basic idea:

- Actors have underlying latent position

Unsupervised Embedding

Basic idea:

- Actors have underlying latent position
- Actors articulate that latent position in their speech

Unsupervised Embedding

Basic idea:

- Actors have underlying latent position
- Actors articulate that latent position in their speech
- This is associated with word usage, so high discriminating words correspond to ideological speech

Unsupervised Embedding

Basic idea:

- Actors have underlying latent position
- Actors articulate that latent position in their speech
- This is associated with word usage, so high discriminating words correspond to ideological speech
- Some words discriminate better than others $\rightsquigarrow$ encode that in our model

Unsupervised Embedding

Basic idea:

- Actors have underlying latent position
- Actors articulate that latent position in their speech
- This is associated with word usage, so high discriminating words correspond to ideological speech
- Some words discriminate better than others $\rightsquigarrow$ encode that in our model

Simplest model: Principal Components

Application of Principal Components in R

Consider press releases from 2005 US Senators

Application of Principal Components in R

Consider press releases from 2005 US Senators
Define $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})$ as the rate senator i uses J words.

Application of Principal Components in R

Consider press releases from 2005 US Senators
Define $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})$ as the rate senator i uses J words.

$$x_{ij} = \frac{\text{No. Times } i \text{ uses word } j}{\text{No. words } i \text{ uses}}$$

Application of Principal Components in R

Consider press releases from 2005 US Senators

Define $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})$ as the rate senator i uses J words.

$$x_{ij} = \frac{\text{No. Times } i \text{ uses word } j}{\text{No. words } i \text{ uses}}$$

`dtm`: 100×2796 matrix containing word rates for senators

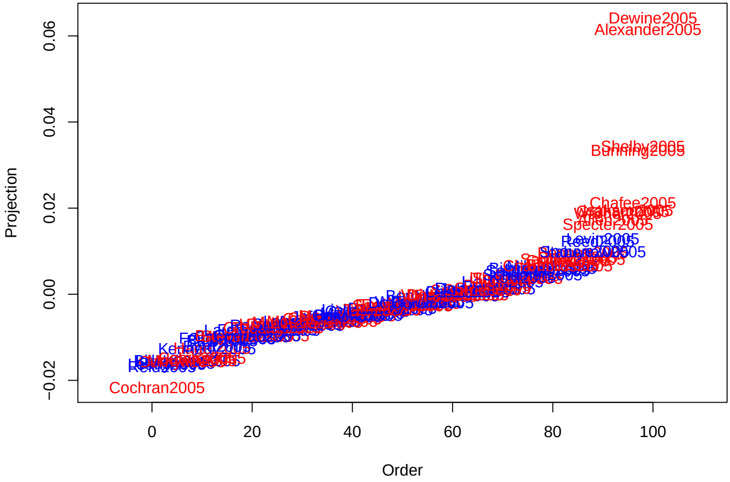
Application of Principal Components in R

Consider press releases from 2005 US Senators
Define $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})$ as the rate senator i uses J words.

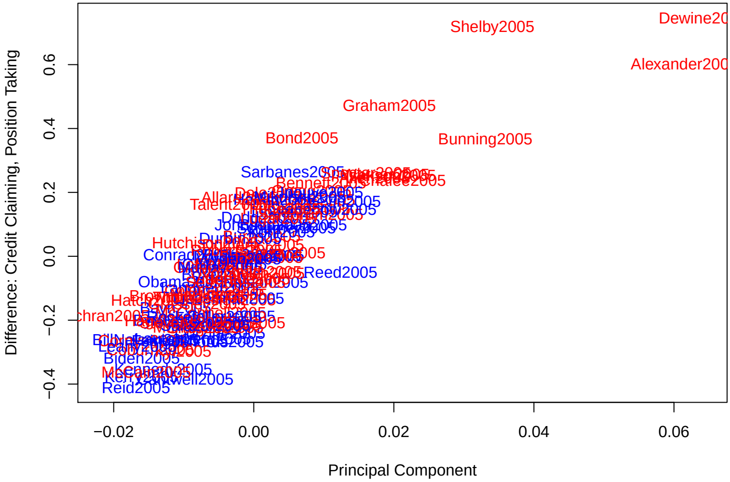
$$x_{ij} = \frac{\text{No. Times } i \text{ uses word } j}{\text{No. words } i \text{ uses}}$$

`dtm`: 100×2796 matrix containing word rates for senators
`prcomp(dtm)` applies principal components

Application of Principal Components in R



Application of Principal Components in R



Probabilistic Unsupervised Embeddings

Principal components is powerful

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models
  c) Tuesday $\rightsquigarrow$ IRT and topic models to scale with votes + text

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models
  c) Tuesday$\rightsquigarrow$ IRT and topic models to scale with votes + text
- Clinton, Jackman, and Rivers (2004)$\rightsquigarrow$ intuition about IRT

Probabilistic Unsupervised Embeddings

Principal components is powerful $\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models
  c) Tuesday $\rightsquigarrow$ IRT and topic models to scale with votes + text
- Clinton, Jackman, and Rivers (2004) $\rightsquigarrow$ intuition about IRT
- Rivers (2002) $\rightsquigarrow$ Identification conditions

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models
  c) Tuesday$\rightsquigarrow$ IRT and topic models to scale with votes + text
- Clinton, Jackman, and Rivers (2004)$\rightsquigarrow$ intuition about IRT
- Rivers (2002)$\rightsquigarrow$ Identification conditions
- Bonica (2014a, 2014b)$\rightsquigarrow$ uses IRT (like the one we're about to use) to scale donors

Probabilistic Unsupervised Embeddings

Principal components is powerful$\rightsquigarrow$ statistical model for unsupervised scaling

Item Response Theory (IRT)

- Origins: educational testing
- Jackman (2002), Clinton, Jackman, and Rivers (2004) apply to roll call voting
- Power of IRT:
  a) Estimate ideal points with few observations
  b) Makes clear how to extend models
  c) Tuesday$\rightsquigarrow$ IRT and topic models to scale with votes + text
- Clinton, Jackman, and Rivers (2004)$\rightsquigarrow$ intuition about IRT
- Rivers (2002)$\rightsquigarrow$ Identification conditions
- Bonica (2014a, 2014b)$\rightsquigarrow$ uses IRT (like the one we're about to use) to scale donors

Monroe and Maeda (2005) and Slopkin and Proksch (2008) develop similar algorithms

WordFish Objective Function

Suppose we have legislator i.

WordFish Objective Function

Suppose we have legislator i.

$$x_{ij} \sim \text{Poisson}(\lambda_{ij})$$

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

Where,

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

Where,

$$
\lambda_{ij} = \text{Rate individual } i \text{ uses word } j
$$

WordFish Objective Function

Suppose we have legislator i.

$$\begin{aligned} x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i) \end{aligned}$$

Where,

$$\begin{aligned} \lambda_{ij} &= \text{Rate individual } i \text{ uses word } j \\ \alpha_i &= \text{Individual } i \text{ loquaciousness} \end{aligned}$$

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

Where,

$$
\begin{aligned}
\lambda_{ij} &= \text{Rate individual } i \text{ uses word } j \\
\alpha_i &= \text{Individual } i \text{ loquaciousness} \\
\psi_j &= \text{Word } j\text{'s frequency}
\end{aligned}
$$

WordFish Objective Function

Suppose we have legislator i.

$$\begin{aligned} x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i) \end{aligned}$$

Where,

$$\begin{aligned} \lambda_{ij} &= \text{Rate individual } i \text{ uses word } j \\ \alpha_i &= \text{Individual } i \text{ loquaciousness} \\ \psi_j &= \text{Word } j\text{'s frequency} \\ \beta_j &= \text{Word } j\text{'s discrimination} \end{aligned}$$

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

Where,

$$
\begin{aligned}
\lambda_{ij} &= \text{Rate individual } i \text{ uses word } j \\
\alpha_i &= \text{Individual } i \text{ loquaciousness} \\
\psi_j &= \text{Word } j\text{'s frequency} \\
\beta_j &= \text{Word } j\text{'s discrimination} \\
\theta_i &= \text{Legislator } i\text{'s latent positions}
\end{aligned}
$$

WordFish Objective Function

Suppose we have legislator i.

$$
\begin{aligned}
x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\lambda_{ij} &= \exp(\alpha_i + \psi_j + \beta_j \times \theta_i)
\end{aligned}
$$

Where,

$$
\begin{aligned}
\lambda_{ij} &= \text{Rate individual } i \text{ uses word } j \\
\alpha_i &= \text{Individual } i \text{ loquaciousness} \\
\psi_j &= \text{Word } j\text{'s frequency} \\
\beta_j &= \text{Word } j\text{'s discrimination} \\
\theta_i &= \text{Legislator } i\text{'s latent positions}
\end{aligned}
$$

"Regression" of x_{ij} on ideal points θ_i, where we have to learn θ_i

WordFish Objective Function/Optimization

Implies the following posterior distribution:

WordFish Objective Function/Optimization

Implies the following posterior distribution:

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) \quad \propto \quad & p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \\
& \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!}
\end{aligned}
$$

WordFish Objective Function/Optimization

Implies the following posterior distribution:

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) \propto p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!}$$

Estimate parameters:

WordFish Objective Function/Optimization

Implies the following posterior distribution:

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) \propto p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!}$$

Estimate parameters:

- EM-algorithm

WordFish Objective Function/Optimization

Implies the following posterior distribution:

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) \propto p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!}$$

Estimate parameters:

- EM-algorithm
- MCMC algorithm

WordFish Objective Function/Optimization

Implies the following posterior distribution:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) &\propto p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \\ & \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!} \end{aligned}$$

Estimate parameters:

- EM-algorithm
- MCMC algorithm
- Variational Approximation

WordFish Objective Function/Optimization

Implies the following posterior distribution:

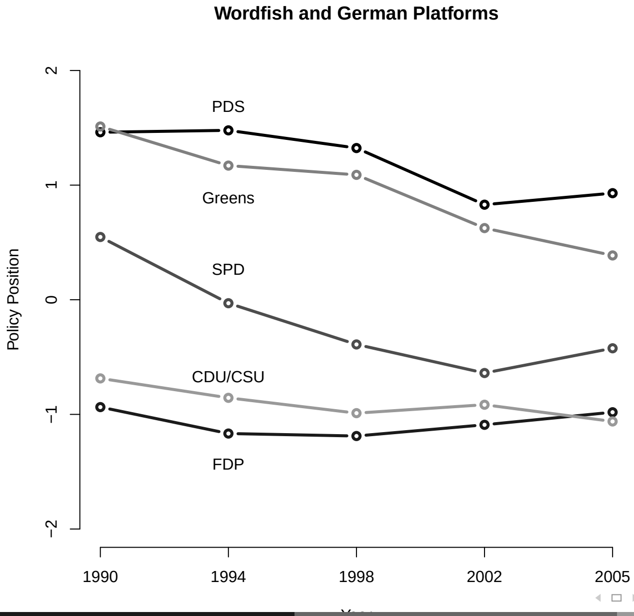
$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta}) \propto p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\psi})p(\boldsymbol{\theta}) \times \prod_{i=1}^{N}\prod_{j=1}^{J} \frac{\exp\left[-\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)\right]\left(\alpha_i + \psi_j + \beta_j \times \theta_i\right)^{x_{ij}}}{x_{ij}!}$$

Estimate parameters:

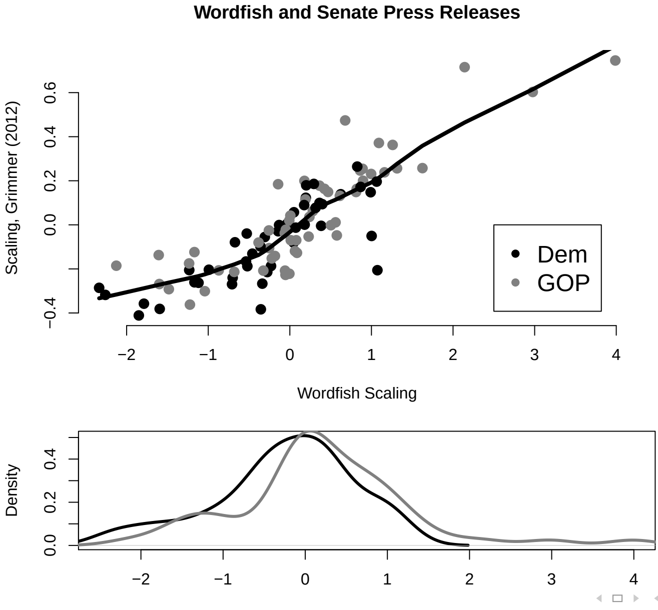
- EM-algorithm
- MCMC algorithm
- Variational Approximation

`Wordfish` package in `R`

Applications: German Party Manifestos



Applications: German Party Manifestos



The Problem with Text-Based Scaling

What does validation mean?

1) Replicate NOMINATE, DIME, or other gold standards?
2) Agreement with experts
3) Prediction of other behavior

Must answer this to make progress on pure text scaling