

# Predicting legislative votes with text

Sean Gerrish and David Blei  
Princeton University Computer Science

1 June 2011

# Predicting Legislative Votes



Y		N	Y		Y
	Y			N	
N	N		Y	N	N
				Y	
				N	

# Predicting Legislative Votes



Y		N	Y		Y
	Y			N	
N	N		Y	N	N
				Y	Y
				Y	Y

# Predicting Legislative Votes



Y		N	Y		Y
	Y			N	
N	N		Y	N	N
		Y		Y	
Y	Y			Y	N

# Predicting Legislative Votes



Y		N	Y		Y
	Y			N	
N	N		Y		N
		Y		Y	
Y	Y		Y	Y	
		N	Y		N

# Predicting Legislative Votes



Y		N	Y		Y
	Y			N	
N	N		Y		Y
		Y		N	
Y	Y		Y	Y	
		N	Y		N
?	?		Y		?

# Organization

## Introduce models

- Overview of a model to represent votes
- Link from documents text to voting model parameters
- Inference on these models

## Experimental results

- Predictive performance
- Text and political sentiment
- Next steps and conclusions

## Ideal points

Ideal points  $x_u$  position lawmakers in a latent political space.

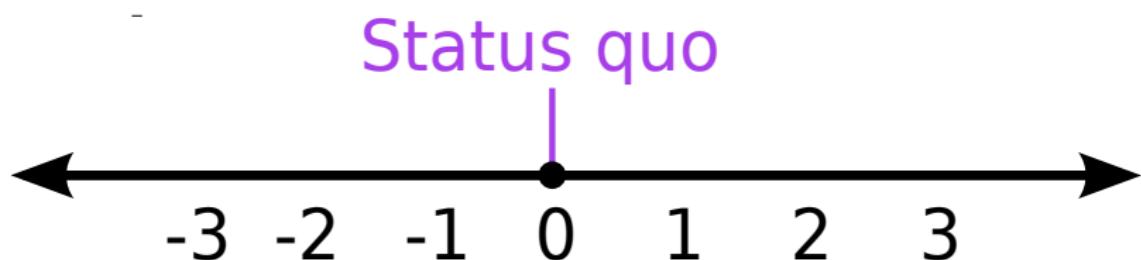
Jackman, 2001

Poole and Rosenthal, 1985

Martin and Quinn, 2002

Jounson and Albert, 1999

Clinton et al., 2004



## Ideal points

Ideal points  $x_u$  position lawmakers in a latent political space.

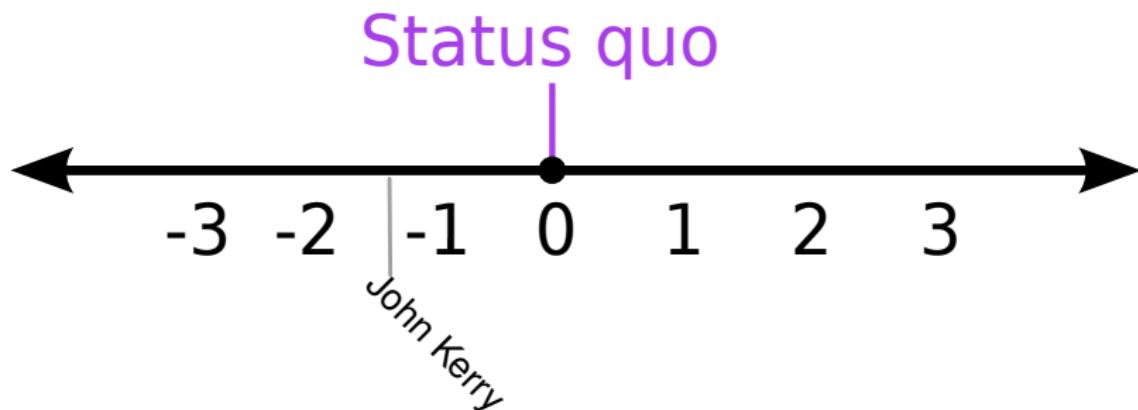
Jackman, 2001

Poole and Rosenthal, 1985

Martin and Quinn, 2002

Jounson and Albert, 1999

Clinton et al., 2004



## Ideal points

Ideal points  $x_u$  position lawmakers in a latent political space.

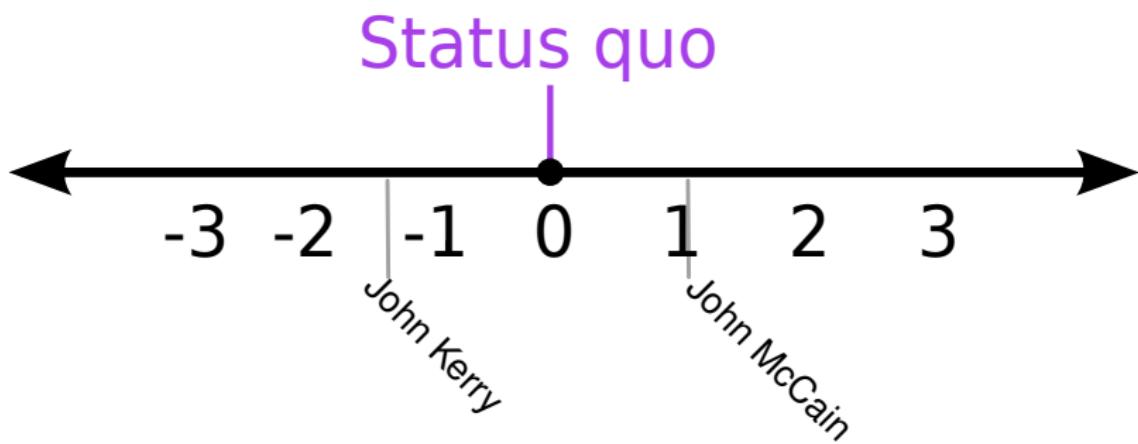
Jackman, 2001

Jounson and Albert, 1999

Poole and Rosenthal, 1985

Clinton et al., 2004

Martin and Quinn, 2002



## Ideal points

Ideal points  $x_u$  position lawmakers in a latent political space.

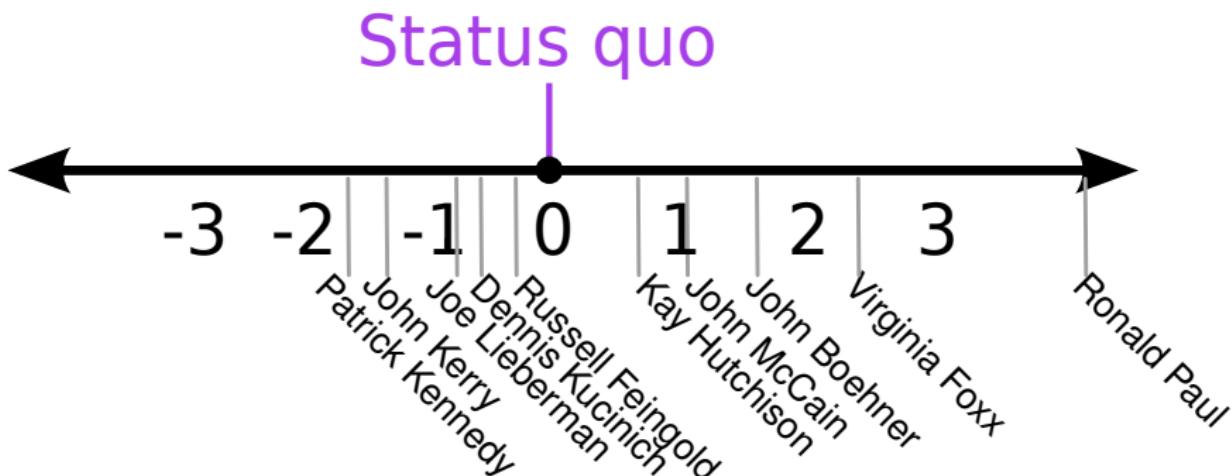
Jackman, 2001

Poole and Rosenthal, 1985

Martin and Quinn, 2002

Jounson and Albert, 1999

Clinton et al., 2004



## Ideal points

Ideal points  $x_u$  position lawmakers in a latent political space.

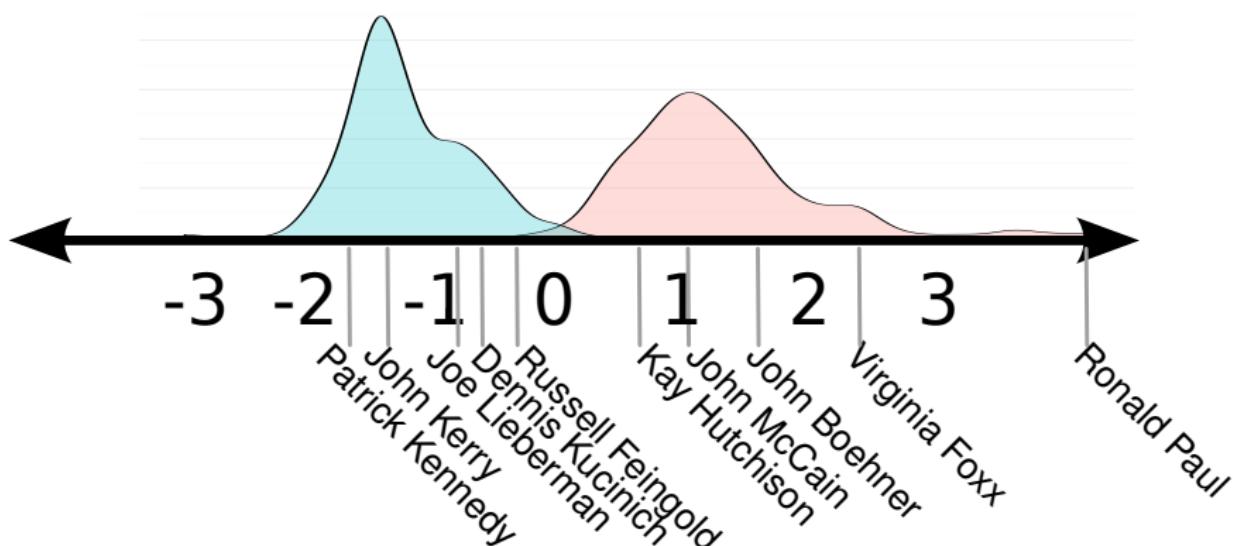
Jackman, 2001

Jounson and Albert, 1999

Poole and Rosenthal, 1985

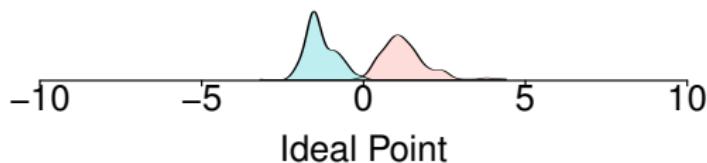
Clinton et al., 2004

Martin and Quinn, 2002



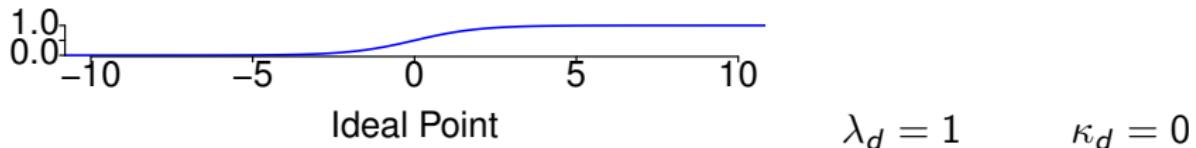
## Documents and ideal points

$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$



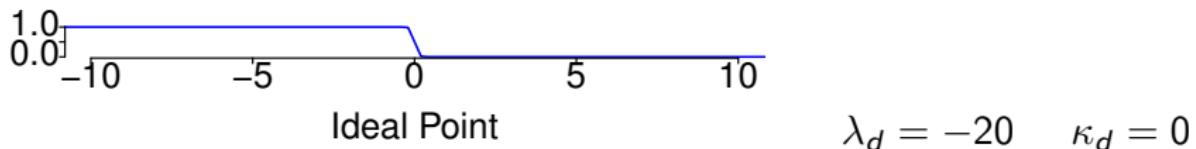
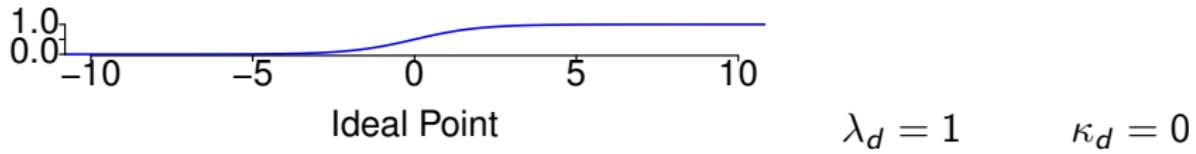
## Documents and ideal points

$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$



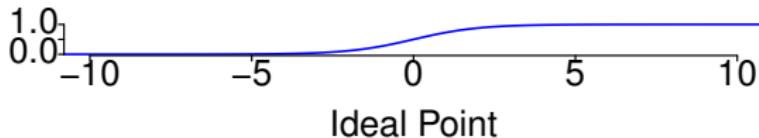
## Documents and ideal points

$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$

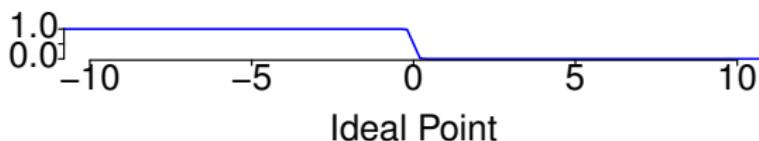


## Documents and ideal points

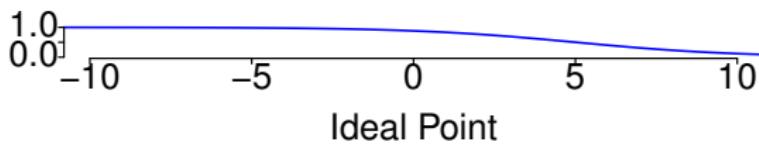
$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$



$$\lambda_d = 1 \quad \kappa_d = 0$$



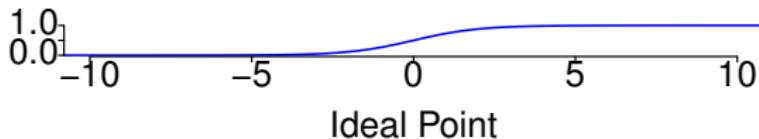
$$\lambda_d = -20 \quad \kappa_d = 0$$



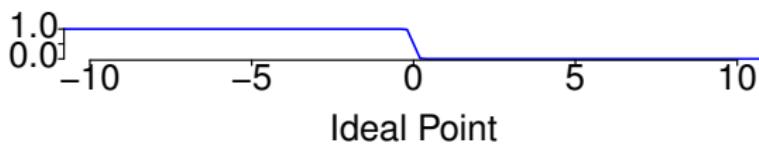
$$\lambda_d = -0.4 \quad \kappa_d = 2$$

## Documents and ideal points

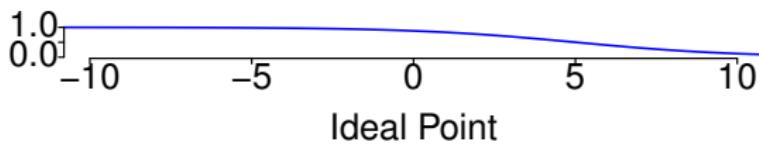
$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$



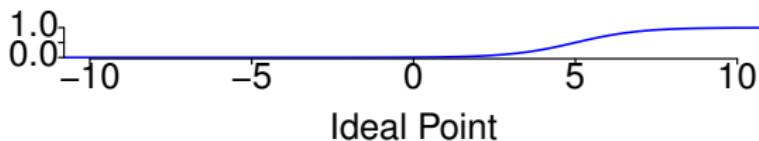
$$\lambda_d = 1 \quad \kappa_d = 0$$



$$\lambda_d = -20 \quad \kappa_d = 0$$



$$\lambda_d = -0.4 \quad \kappa_d = 2$$



$$\lambda_d = 1 \quad \kappa_d = -5$$

# The Ideal Point Model

Jackman, 2001

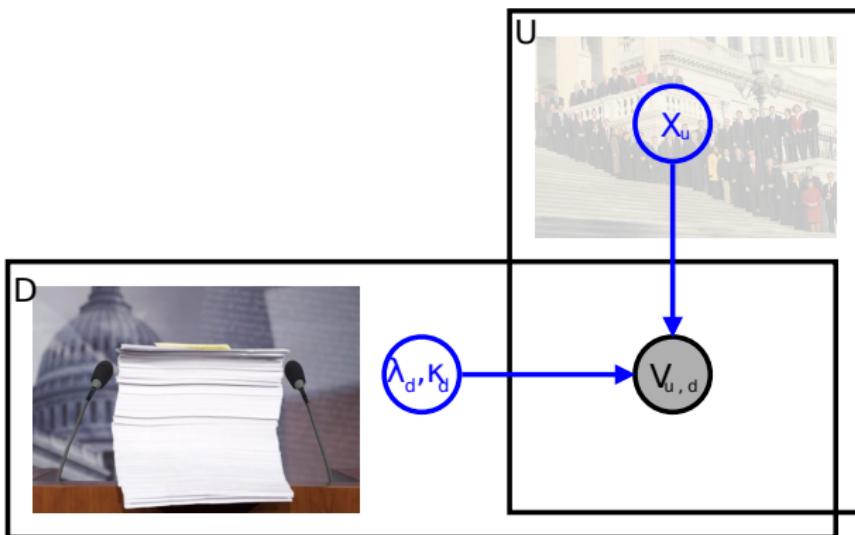
Poole and Rosenthal, 1985

Martin and Quinn, 2002

Jounson and Albert, 1999

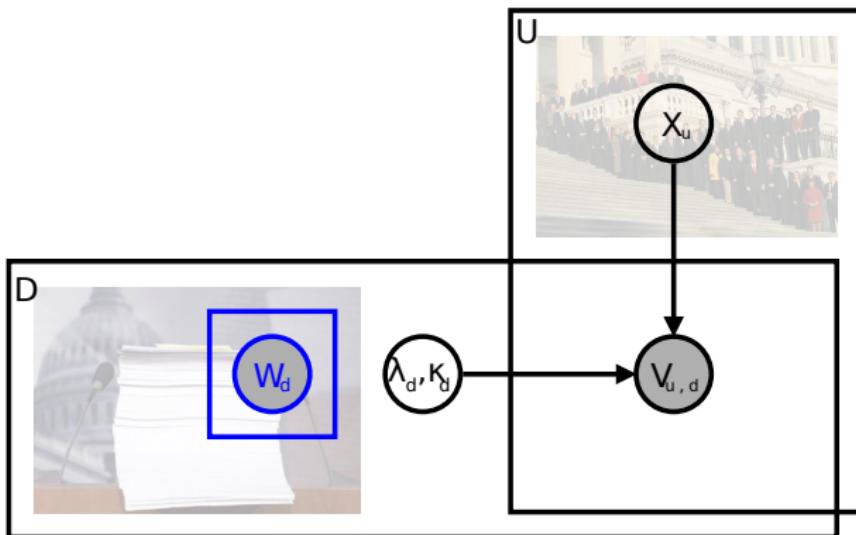
Clinton et al., 2004

$$p(v_{ud} = \text{Yes} | x_u, \lambda_d, \kappa_d) = \text{logistic}(x_u \cdot \lambda_d + \kappa_d)$$



## Adding text to the ideal point model

We can use the text of legislation to infer bills' positions.

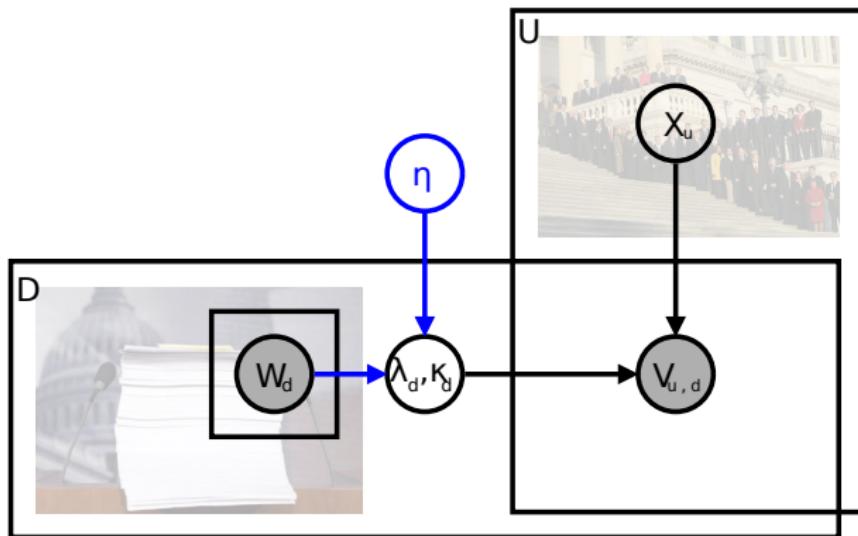


# Ideal Point Text Regression

Regress document parameters on word counts:

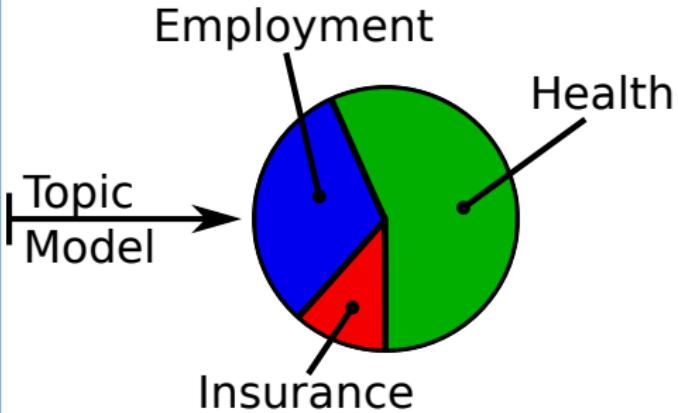
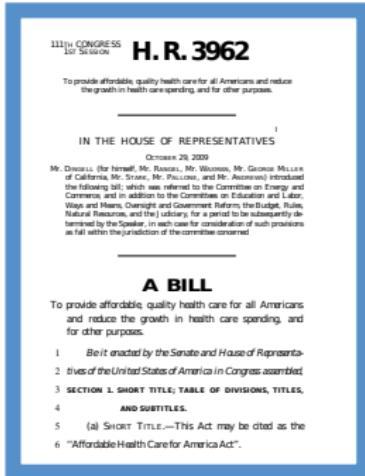
$$\lambda_d \sim \mathcal{N}(\boldsymbol{\eta}_\lambda^T \mathbf{w}_d, \sigma^2)$$

$$\kappa_d \sim \mathcal{N}(\boldsymbol{\eta}_\kappa^T \mathbf{w}_d, \sigma^2)$$



## Ideal Point Topic modeling

Use supervised topics to infer document parameters. Blei et al. 2008

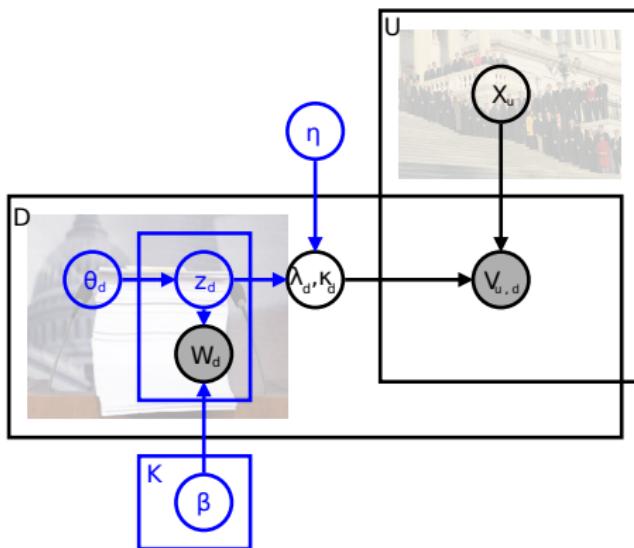


# Ideal Point Topic modeling

Use supervised topics to infer document parameters. Blei et al. 2008

$$\lambda_d \sim \mathcal{N}(\boldsymbol{\eta}_\lambda^T \bar{\mathbf{z}}_d, \sigma^2)$$

$$\kappa_d \sim \mathcal{N}(\boldsymbol{\eta}_\kappa^T \bar{\mathbf{z}}_d, \sigma^2)$$



## Posterior Inference

Recall that we only observe words  $\mathbf{W}$  and votes  $V$ .  
We are interested in the posterior

$$p(\lambda_d, \kappa_d, x_u, \eta | V, \mathbf{W})$$

We derived a mean-field variational inference algorithm.

- This involves positing a family of fully factorized posterior distributions and finding the distribution from this family which is “closest” in K-L divergence to the true posterior.
- The resulting ideal points are correlated at over 0.98 with MCMC ideal points (the standard in this field).
- Variational methods like this are amenable to inference in large-scale datasets. Hoffman et al. 2010

## Derivation and implementation of variational inference

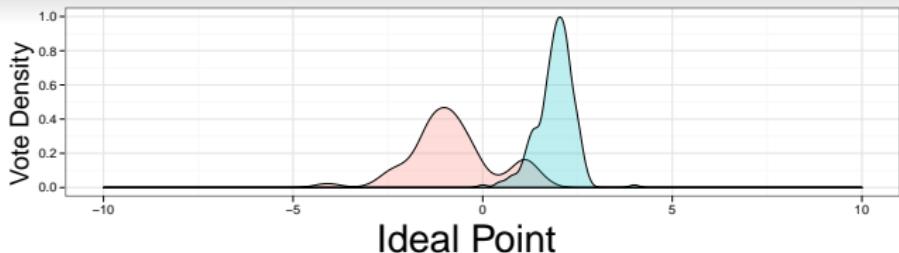
Finding the variational posterior involves optimizing a lower bound on the model evidence  $p(V, \mathbf{W})$ .

This objective is optimized via gradient ascent, and we can use a few tricks to find the objective and converge better, e.g.:

- Use a second-order delta approximation for the bound Bickel et al. 2007
- Update subsets of lawmakers and legislation in a round-robin fashion to avoid cycles

## Experiments

- We fit this model to each congressional session (2-year period) from Jan 1997 to Jan 2011.
- Data is from Govtrack.us
- There were:
  - 4,447 bills
  - 1,269 lawmakers
  - 1,837,033 yea/nay roll-call votes



11TH CONGRESS  
1st Session

## H.R. 3962

To provide affordable, quality health care for all Americans and reduce the growth in health care spending, and for other purposes.

---

### IN THE HOUSE OF REPRESENTATIVES

October 29, 2009  
Mr. DeLise, the Honorable John C. Waldron, Mr. George Miller  
of California, Mr. Platts, Mr. Pallotta, and Mr. Aspinwall introduced  
the following bill; which was referred to the Committee on Energy and  
Commerce, and in addition to the Committees on Education and Labor,  
Wayne L. Joesten, Chairman; on Small Business; on Science, Space,  
Natural Resources, and the Judiciary; for a period to be subsequently de-  
termined by the Speaker, in each case for consideration of such provisions  
as the author may prescribe for the committee concerned

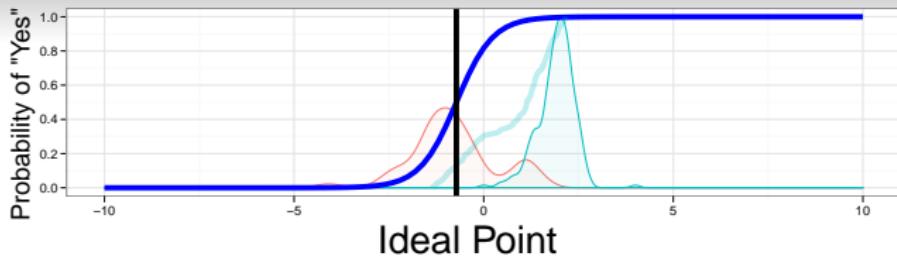
---

### A BILL

To provide affordable, quality health care for all Americans  
and reduce the growth in health care spending, and  
for other purposes.

1. Be it enacted by the Senate and House of Representa-
2. tives of the United States of America in Congress assembled,
3. SECTION 1. SHORT TITLE; TABLE OF CONTENTS; TITLES;
4. AND SUBTITLES.
5. (a) SHORT TITLE.—This Act may be cited as the
6. “Affordable Health Care for America Act”.

- 276 of 311 Democrats voted “Yes”.
- 217 of 217 Republicans voted “No”.



Digitized by srujanika@gmail.com  
Digitized by srujanika@gmail.com

To provide affordable, quality health care for all Americans and reduce the growth in health care spending, and for other purposes.

## IN THE HOUSE OF REPRESENTATIVES

Orvalmy, 20-208

Mr. DONALD, Mr. KENNEDY, Mr. WADDELL, Mr. GLENISTER McLELLAN of California, Mr. WEINER, Mr. TAYLOR, and Mr. ASWIDGE introduced the following bill, which was referred to the Committee on Energy and Commerce, and in addition to the Committee on Education and Labor, Ways and Means, Oversight and Government Reform, the Budget, Ethics, Natural Resources, and the Judiciary, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned:

A BILL.

To provide affordable, quality health care for all Americans and reduce the growth in health care spending, and

Be it enacted by the Senate and House of Representatives  
of the United States of America in Congress assembled,

3 SECTION 1. SHORT TITLE; TABLE OF DIVISIONS, TITLES,

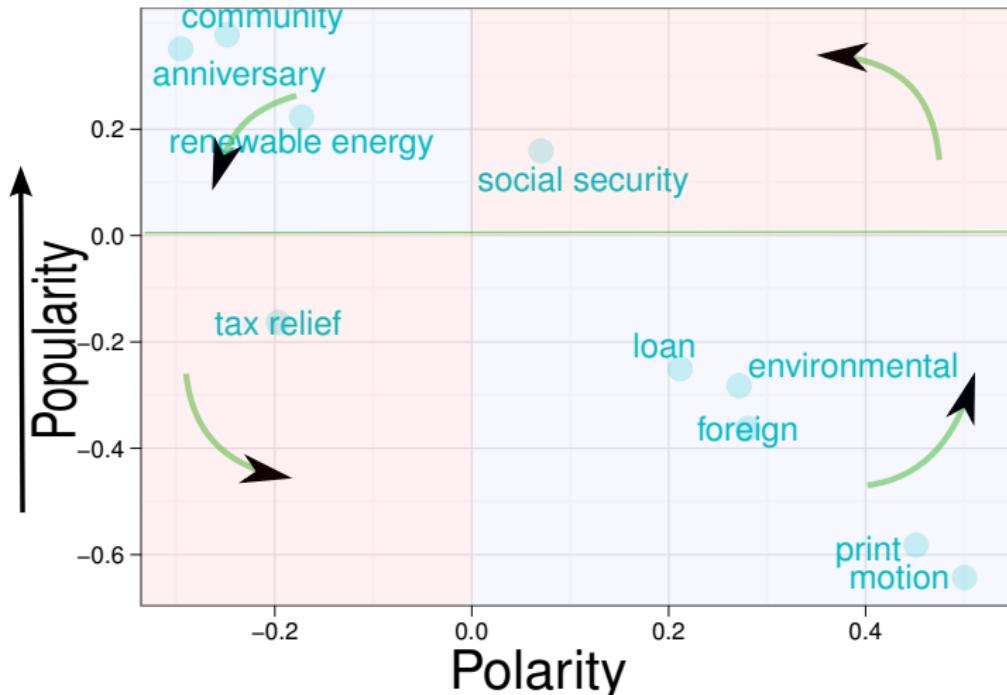
(a) SHORT TITLE.—This Act may be cited as the "Affordable Health Care for America Act".

- 276 of 311 Democrats voted “Yes”.
  - 276 of 276 Republicans voted “No”.
  - Primary topic:  
*care, applicable, coverage, hospital, eligible*
  - Inferred bill parameters are

$$\lambda_d = 1.5, \kappa_d = 2.1$$

- Accuracy: 79%
  - Baseline accuracy: 51%

## Example - Ridge Regression parameters $\eta_\lambda, \eta_\kappa$



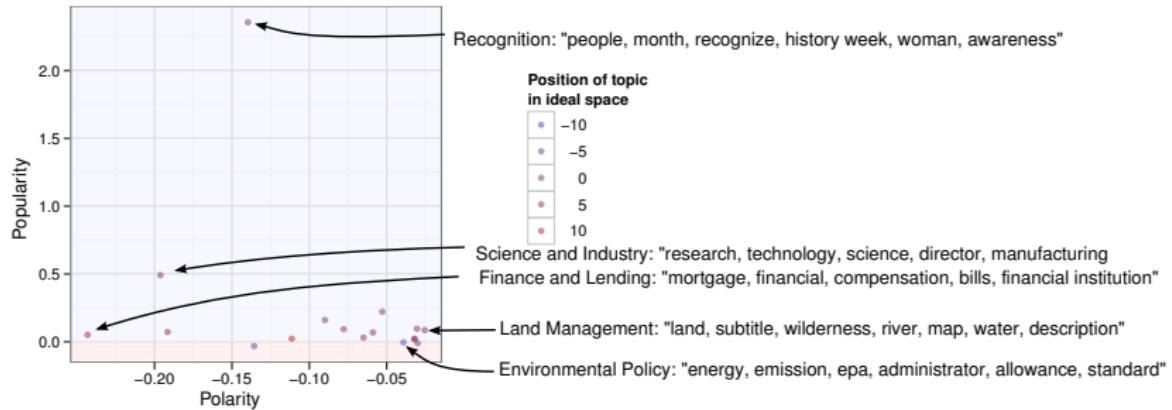
ideal-point text regression, the most popular terms are “life”.

With

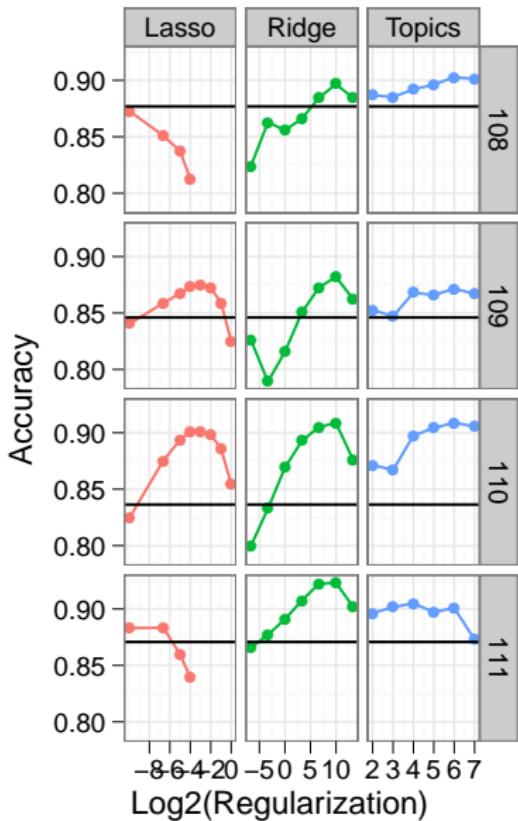
# Results - Topics

Description	Topic summary (Example words)
Popular (high $\eta_\kappa$ )	Military recognition (veteran, war, serve, veterans, training, military) Recognition (people, month, recognize, history, week, woman)
Unpopular	Anything polarizing
Left-wing (low $-\frac{\eta_\kappa}{\eta_\lambda}$ )	Employment and eligibility (eligible, credit, head, qualified, training) Environmental policy (energy, emission, epa, administrator, allowance) Regulation (transfer, requires, contract, transportation, expense, corporation)
*Right-wing (high $-\frac{\eta_\kappa}{\eta_\lambda}$ )	Social Services (medicare, social security act, clause, hospital) Land management (land, subtitle, wilderness, river) Tobacco (delivery, cigarette, tobacco, smokeless, sale, bills)

# Results - Topics



## Results - Prediction



- Predictive accuracy as high as 92% with these models
- Works best for 32-64 topics
- Baseline “assume everyone votes yes” is black horizontal line
- 93% correct votes with a little more effort

## Current and future directions

Model ideal points over time, or with Kalman-type updates

Multiple ideal dimensions

Apply these models to other datasets:

- Content recommendation
- Keyword-based advertising
- Music recommendation

Use extra information:

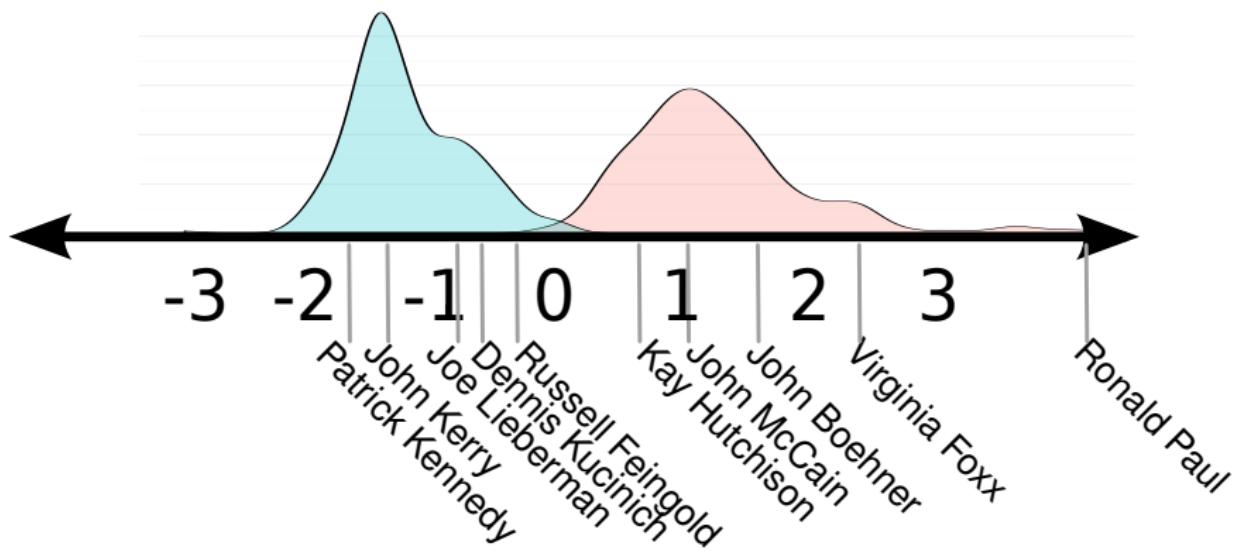
- Transcripts of floor debates
- Per-topic ideal points (uses more than 1 dimension)
- Model ideal points over time

# Summary

The models we have discussed...

- Are predictive models for lawmakers' votes,
- Include an exploratory model for discovering politicized topics and understanding the themes driving lawmakers, and
- Have two key ideas at heart,
  - Classic ideal point models
  - Supervision with text (e.g., topic models)
- Can be used in a general collaborative filtering setting.

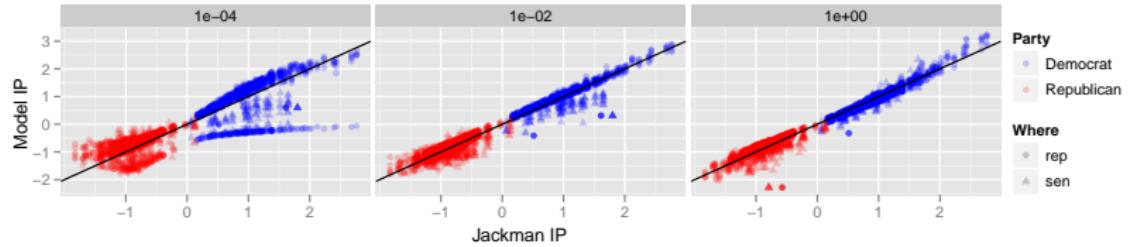
## Contact



## Contact

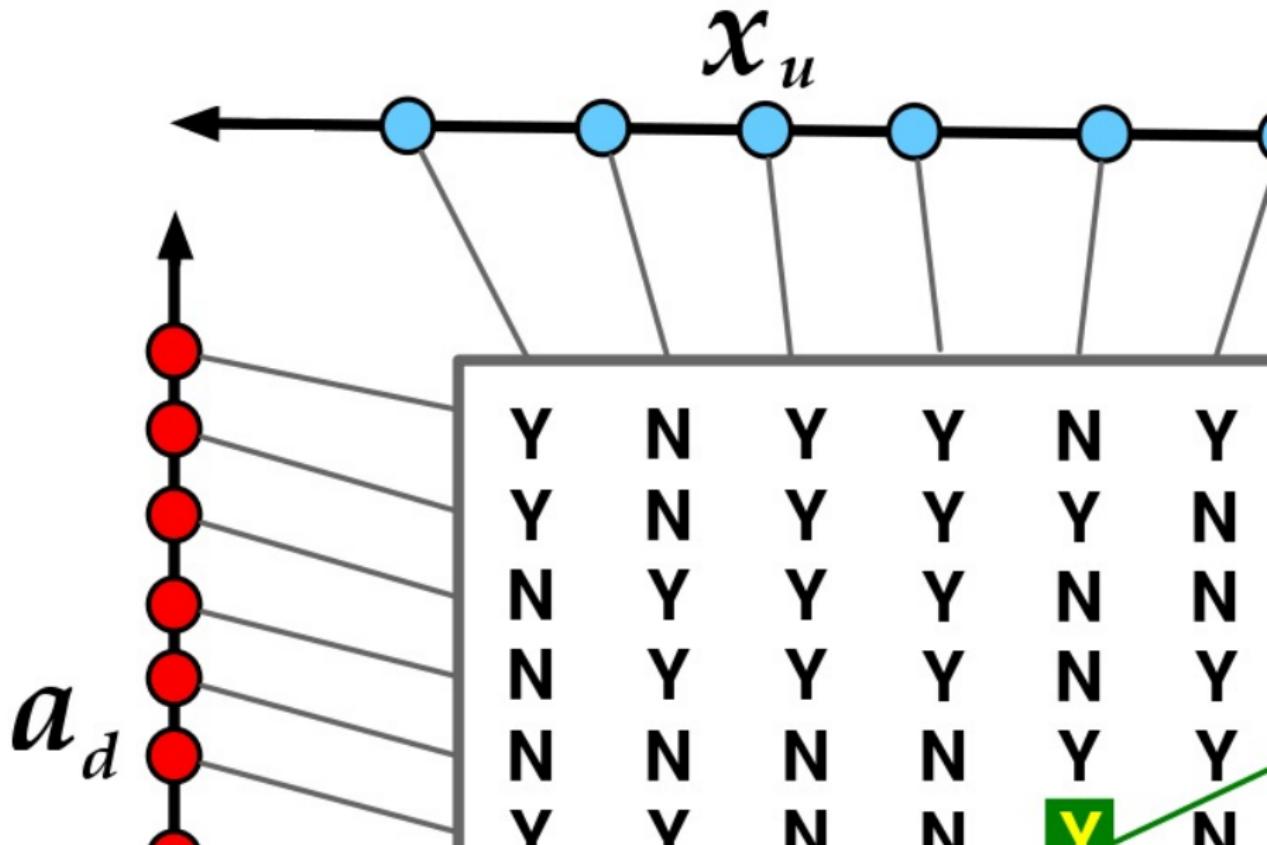
- Sean Gerrish ([sgerrish@cs.princeton.edu](mailto:sgerrish@cs.princeton.edu))
- David Blei ([blei@cs.princeton.edu](mailto:blei@cs.princeton.edu))

# Appendix

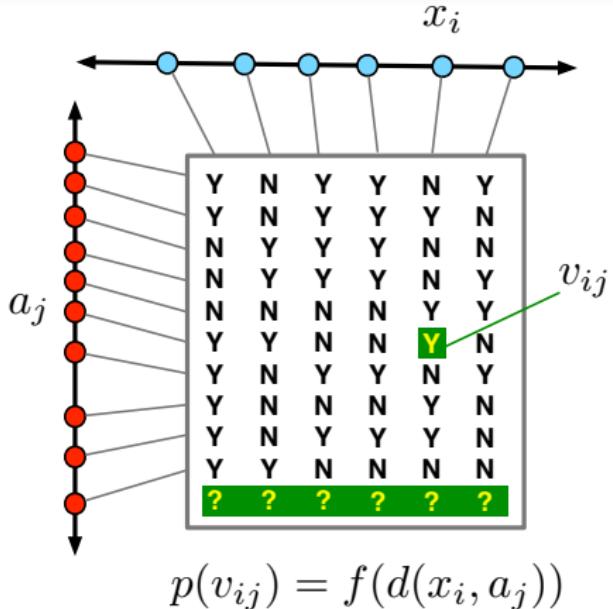


# Ideal Point Topic Models

## The ideal point model



## The ideal point model is limited for prediction

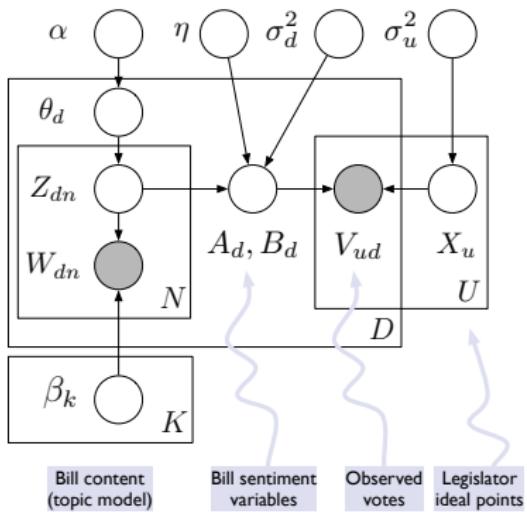


- We can predict a missing vote.
- But we cannot predict all the missing votes from a bill.
- Cf. the limitations of collaborative filtering

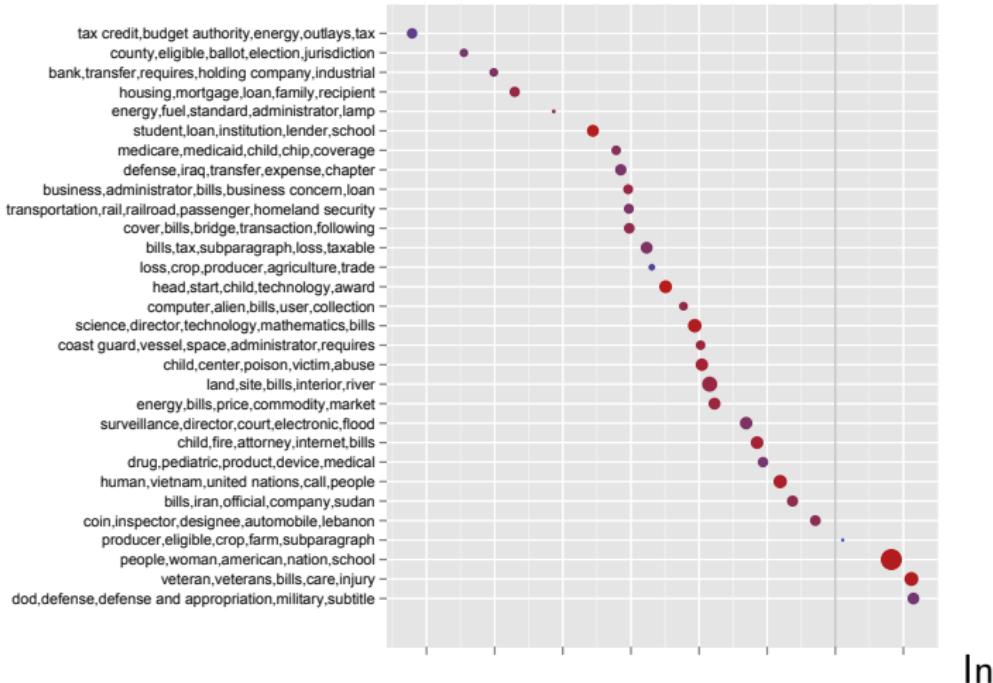
## Ideal point topic models

Use supervised topic modeling assumptions as a predictive mechanism from bill texts to bill discrimination.

# Ideal point topic models

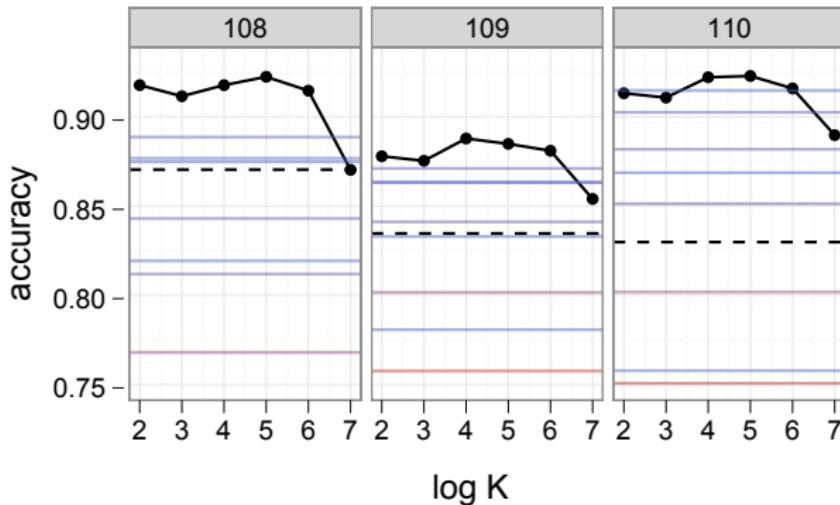


# Ideal point topics



In addition to senators and bills, IPTM places **topics** on the spectrum.

## Prediction on completely held-out votes



Versus the LASSO, the IPTM correctly predicted 126,000 more votes.

## Ideal point topic models

- Ideal point topic model illustrates
  - Topic modeling embedded in a complex model
  - Topic modeling used to solve a real-world problem with text
- More generally, consider collaborative filtering.
  - Senators are *users*.
  - Bills are *items*.
- Existing collaborative filtering is akin to classical ideal point.
- Our model lets us predict preferences on *completely new items*.