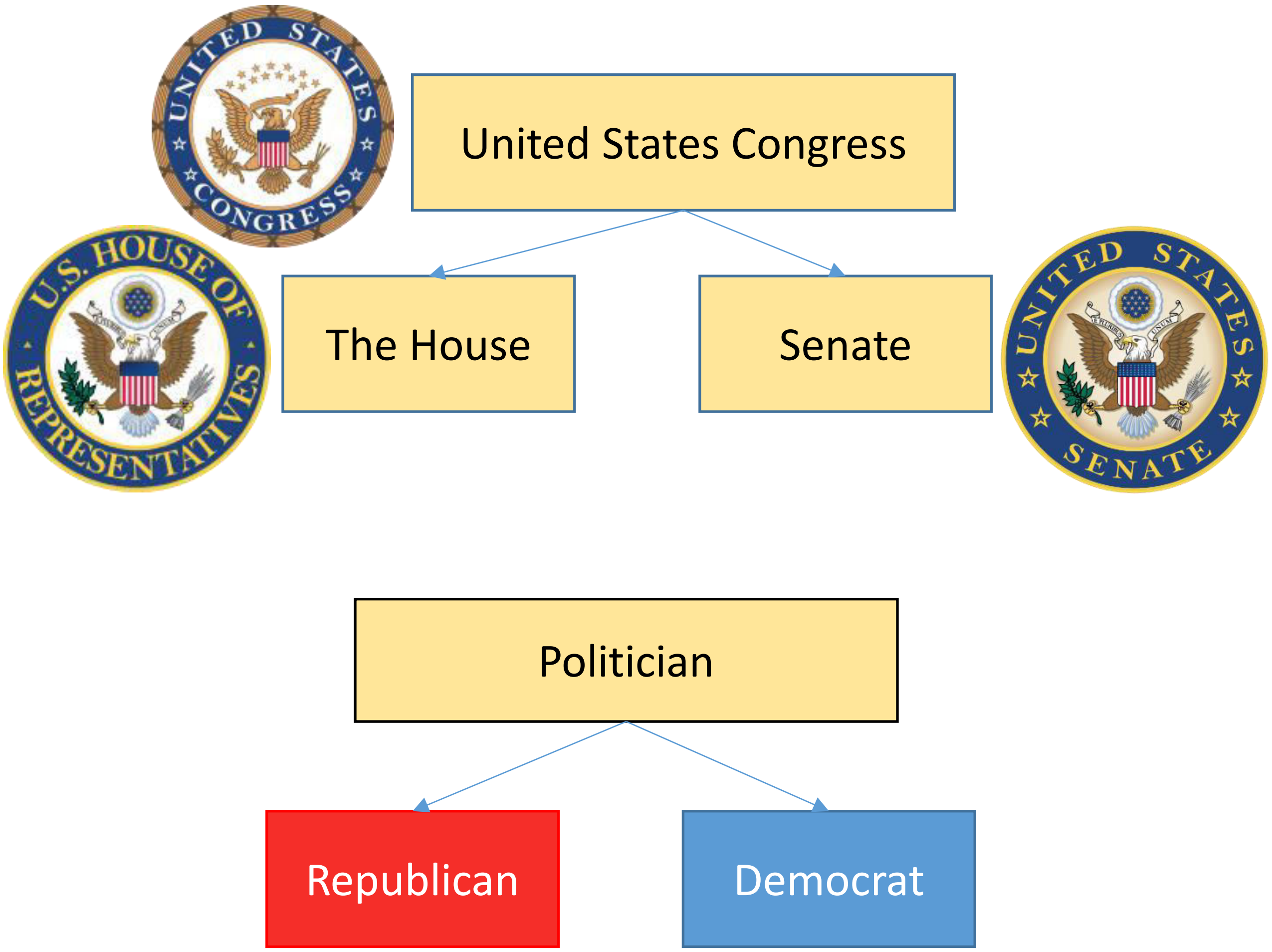
Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network

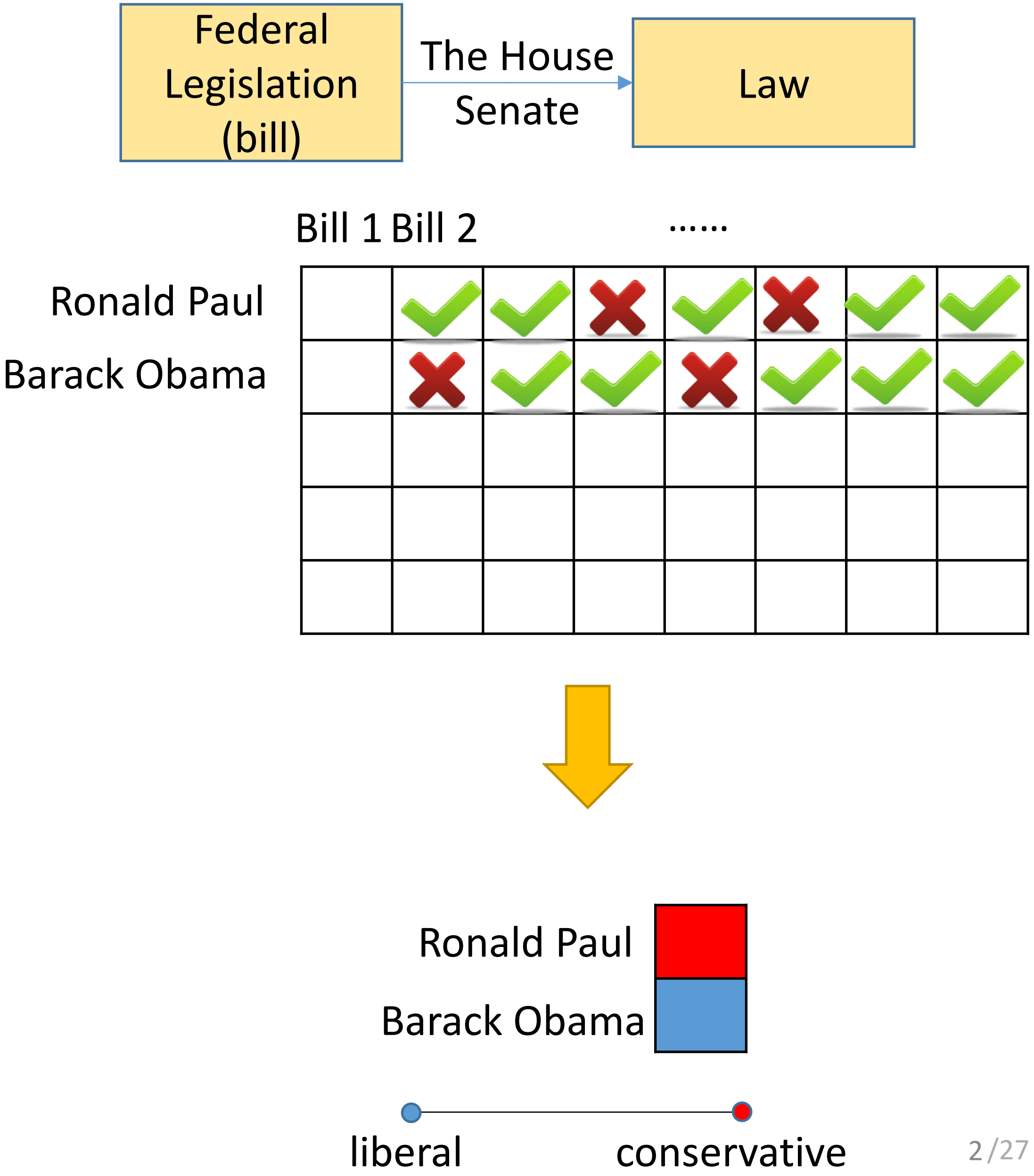
Yupeng Gu[†], Yizhou Sun[†], Ning Jiang[‡], Bingyu Wang[†], Ting Chen[†]

[†]Northeastern University

[‡]University of Illinois at Urbana-Champaign

Background

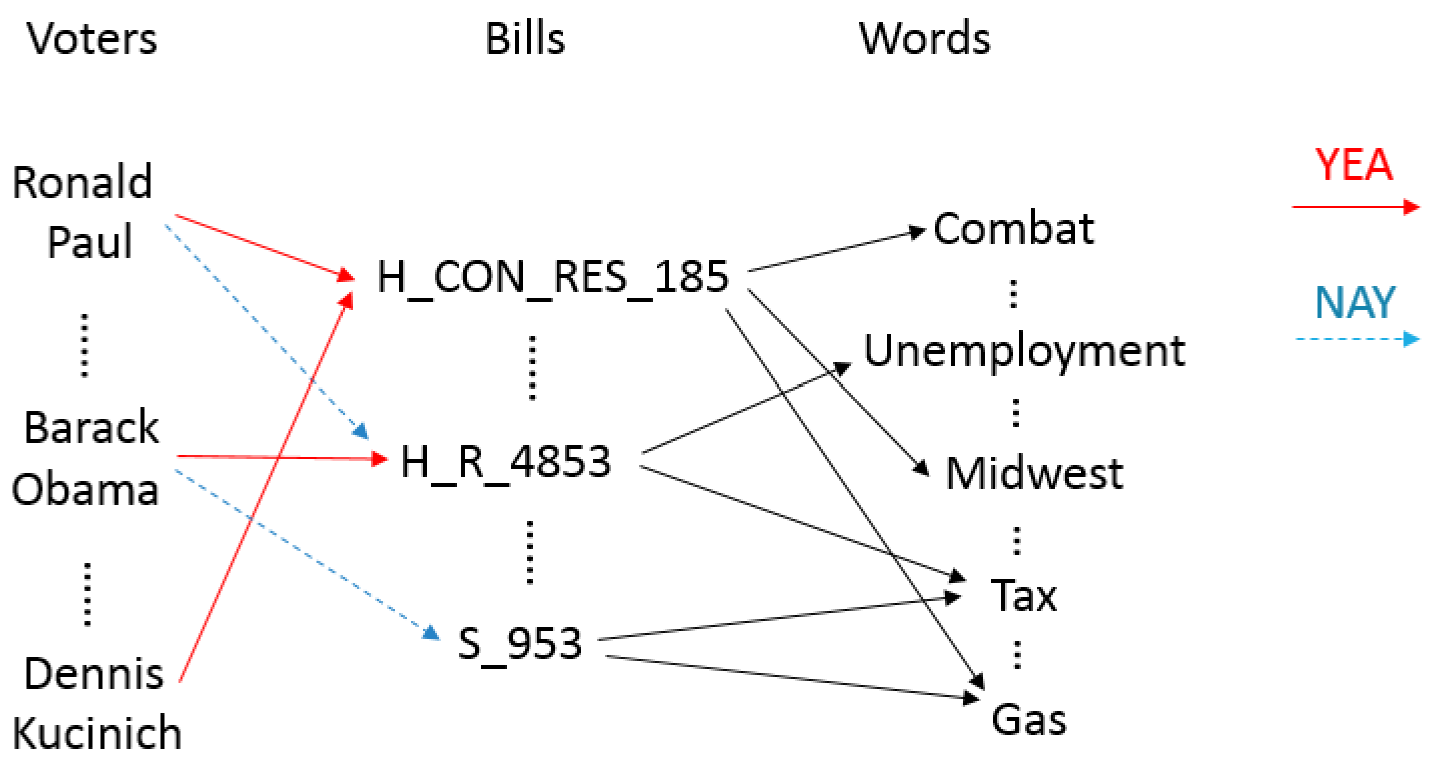

United States Congress
The House
Senate
Politician
Republican
Democrat


Federal Legislation (bill)
The House
Senate
Law
Bill 1 Bill 2
……
Ronald Paul
Barack Obama
Ronald Paul
Barack Obama
liberal
conservative

Outline

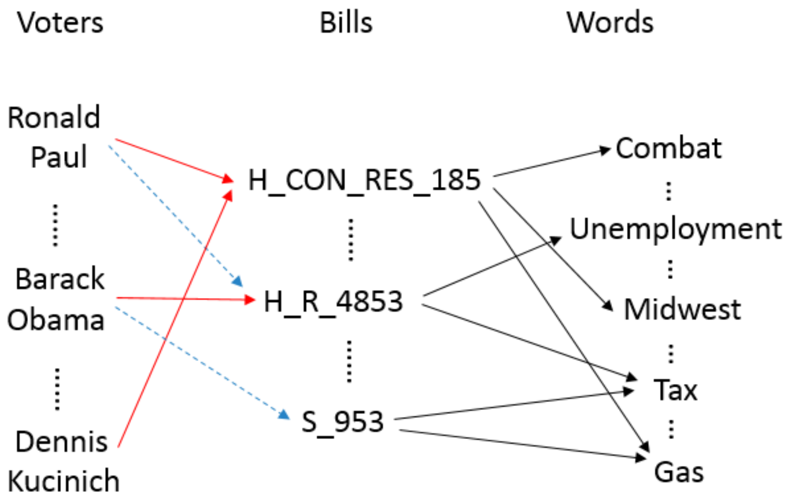
- **Background and Problem Definition**
- Topic Factorized Ideal Point Model and Learning Algorithm
- Experimental Results
- Conclusion

Legislative Voting Network

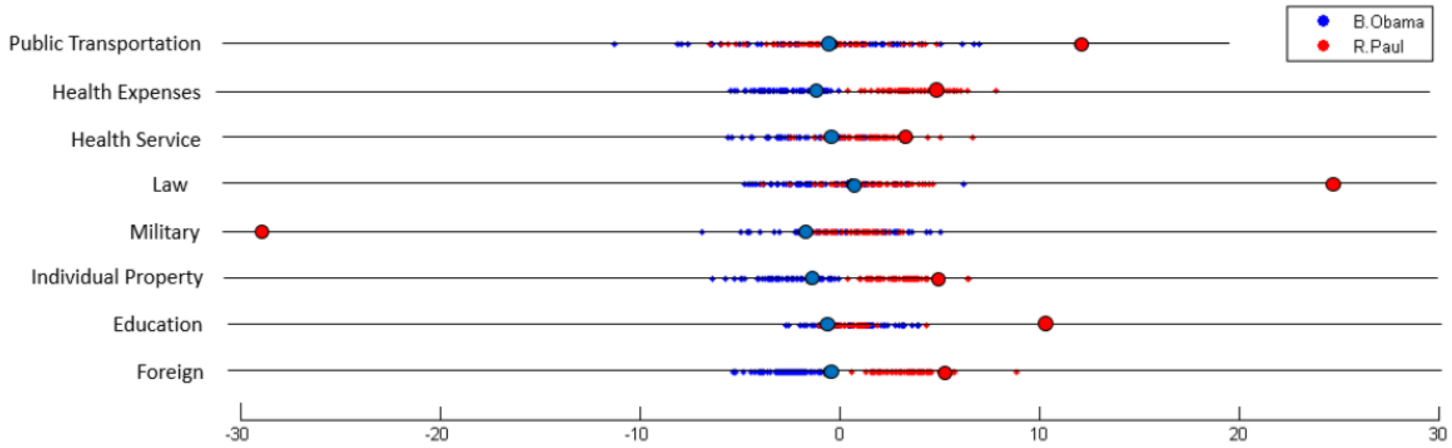


Problem Definition

Input:
Legislative Network



Output:
$\boldsymbol{x}_u$: Ideal Points for Politician u
$\boldsymbol{a}_d$: Ideal Points for Bill d



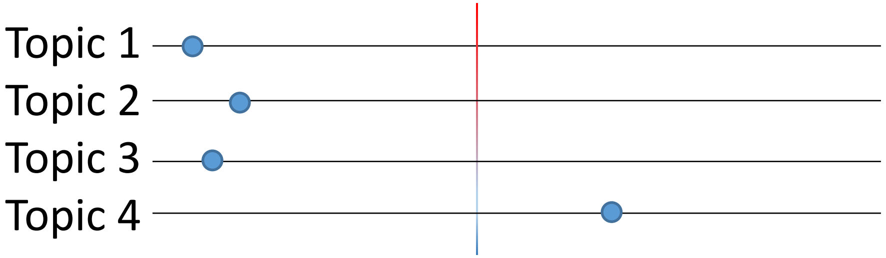
$\boldsymbol{x}_u$'s on different topics

Existing Work

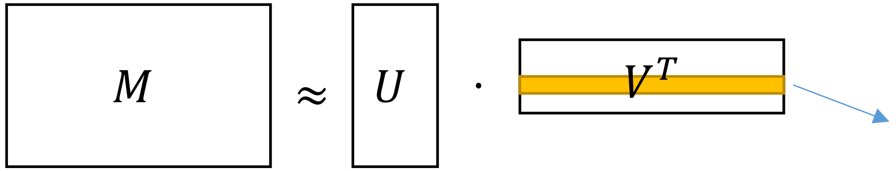
- 1 dimensional ideal point model (Poole and Rosenthal, 1985; Gerrish and Blei, 2011)
- High-dimensional ideal point model (Poole and Rosenthal, 1997)
- Issue-adjusted ideal point model (Gerrish and Blei, 2012)

Motivations

- Voters have different attitudes on different topics.



- Traditional matrix factorization method cannot give the meanings for each dimension.



k^{th} latent factor

- Topics of bills can influence politician's voting, and the voting behavior can better interpret the topics of bills as well.

Topic Model:

- Health
- Public Transport
- …

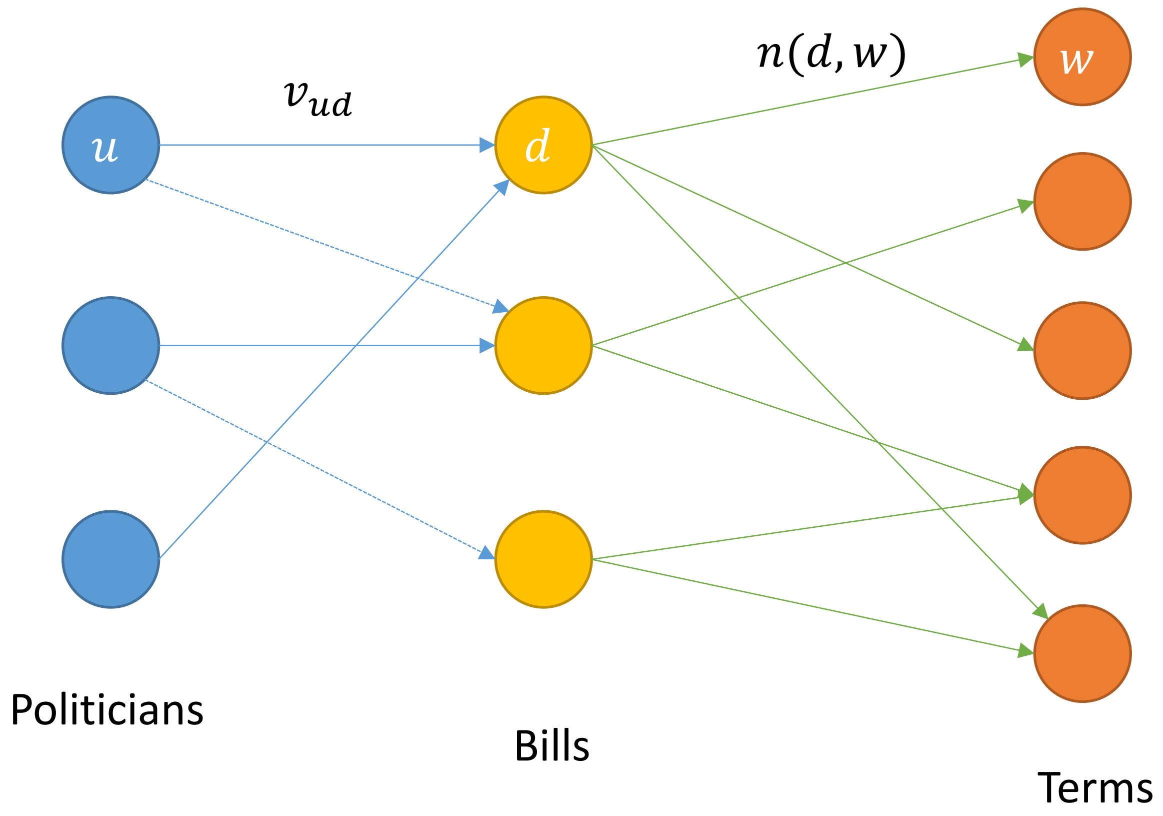
Voting-guided Topic Model:

- Health Service
- Health Expenses
- Public Transport
- …

Outline

- Background and Problem Definition
- **Topic Factorized Ideal Point Model and Learning Algorithm**
- Experimental Results
- Conclusion

Topic Factorized IPM

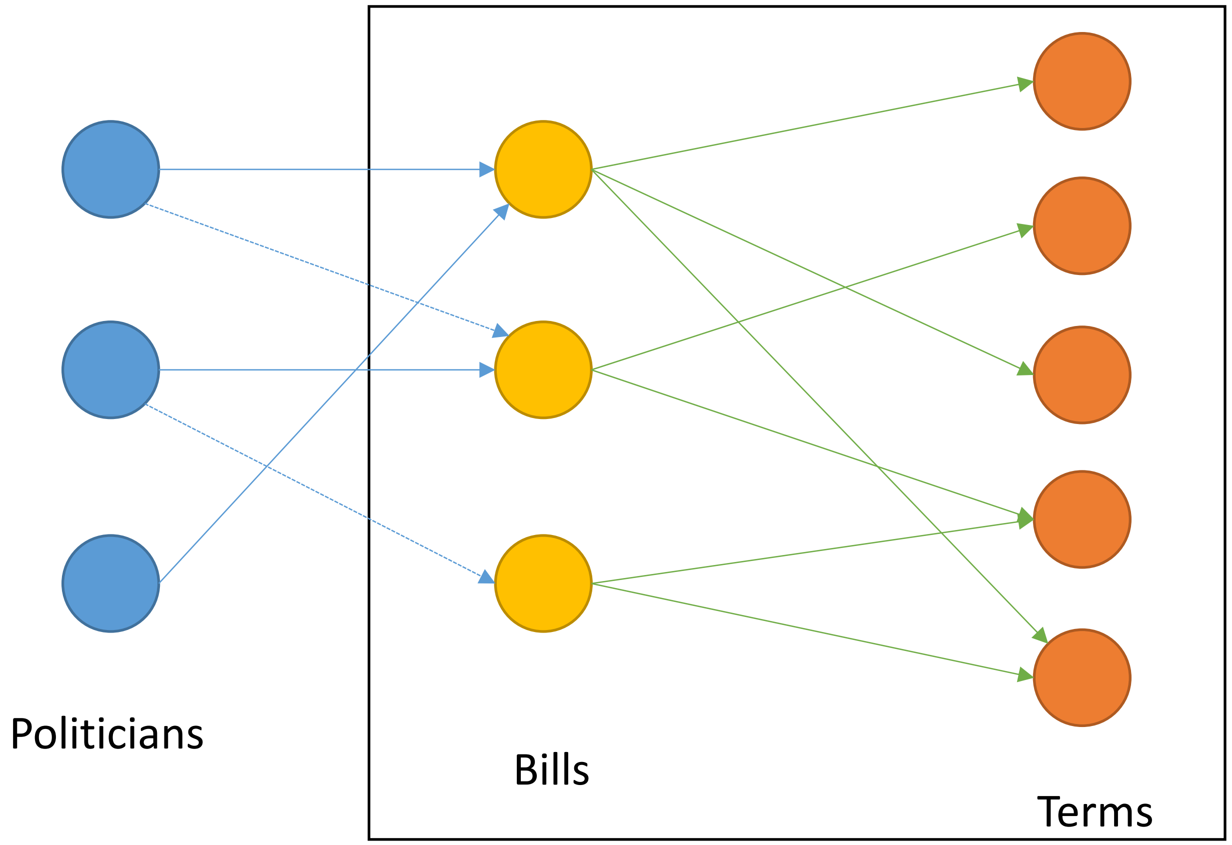


Heterogeneous Voting Network

Entities:
- Politicians
- Bills
- Terms

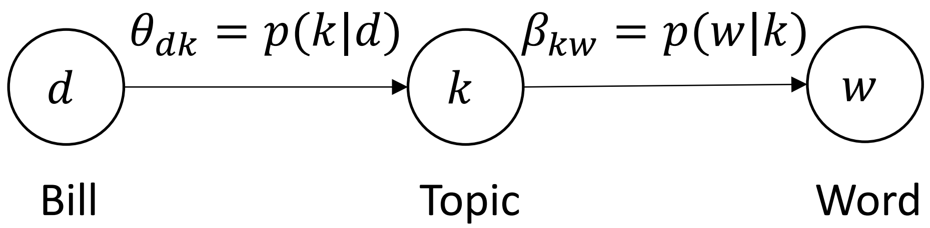
Links:
- (P, B)
- (B, T)

Text Part



Text Part

- We model the probability of each word in each document as a mixture of multinomial distributions, as in PLSA (Hofmann, 1999) and LDA (Blei et al., 2003)

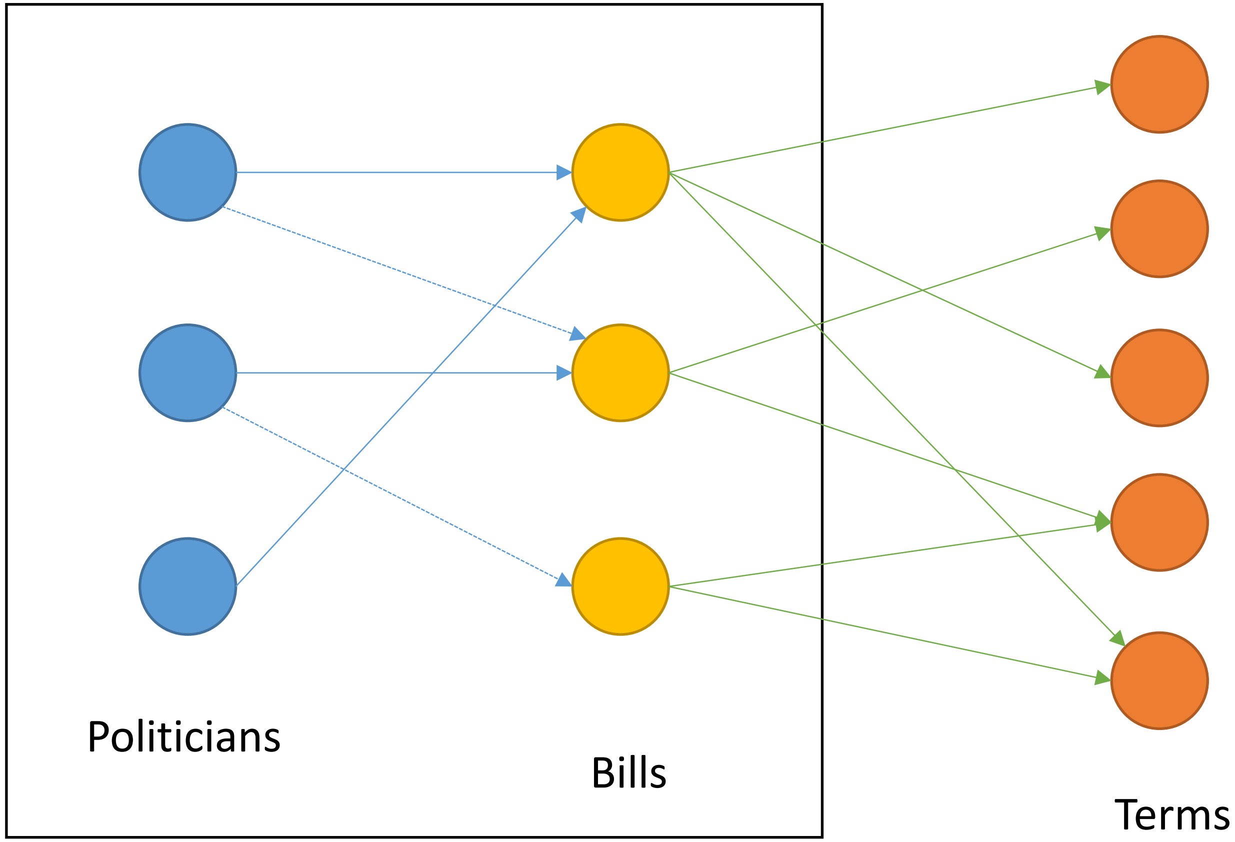


$$\boldsymbol{w}_d = \big(n(d,1), n(d,2), \dots, n(d,N_w)\big)$$

$$p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_w \Big(\sum_k \theta_{dk}\beta_{kw}\Big)^{n(d,w)}$$

$$p(\boldsymbol{W}|\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_d \prod_w \Big(\sum_k \theta_{dk}\beta_{kw}\Big)^{n(d,w)}$$

Voting Part



Voting Part

	Topic 1	Topic 2	......		Topic k	......	Topic K	
Voter u $\boldsymbol{x}_u$	x_{u1}	x_{u2}			x_{uk}		x_{uK}	$x_{uk} \in \boldsymbol{R}$
Bill d $\boldsymbol{a}_d$	a_{d1}	a_{d2}			a_{dk}		a_{dK}	$a_{dk} \in \boldsymbol{R}$

	d_1	d_2		......				d_{N_D}
u_1	0	1	-1	1	1	1	1	1
u_2	0	0	-1	1	1	1	-1	1
⋮								
u_{N_U}	1	1	1	1	-1	1	0	0

User-Bill voting matrix $\boldsymbol{V}$

θ_{d1} x_{u1} a_{d1}

⋮

θ_{dk} x_{uk} a_{dk}

⋮

θ_{dK} x_{uK} a_{dK}

$$\cancel{\hat{r}_{ud} = \sum_{k=1}^{K} x_{uk} a_{dk}}$$

$$\cancel{\hat{r}_{ud} = \sum_{k=1}^{K} \theta_{dk} x_{uk} a_{dk}}$$

$$p(v_{ud} = 1) = \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d) \xrightarrow{\text{YEA}}$$

$$p(v_{ud} = -1) = 1 - \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d) \xrightarrow{\text{NAY}}$$

$$p(\boldsymbol{V}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = \prod_{(u,d):v_{ud}\neq 0} (p(v_{ud} = 1)^{\overset{I_{\{v_{ud}=1\}}}{\boxed{\frac{1+v_{ud}}{2}}}} p(v_{ud} = -1)^{\overset{I_{\{v_{ud}=-1\}}}{\boxed{\frac{1-v_{ud}}{2}}}})$$

Combining Two Parts Together

- The final objective function is a linear combination of the two average log-likelihood functions over the *word links* and *voting links*.

$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1-\lambda)\boxed{\frac{\sum_{d,w} n(d,w)\log(\sum_k \theta_{dk}\beta_{kw})}{N_F}} + \lambda\boxed{\frac{\sum_{(u,d):v_{ud}\neq 0}(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1))}{N_V}}$$

s.t.

$$0 \le \theta_{dk} \le 1, \quad \sum_k \theta_{dk} = 1 \quad \text{and} \quad 0 \le \beta_{kw} \le 1, \quad \sum_w \beta_{kw} = 1$$

- We also add an l_2 regularization term to A and X to reduce over-fitting.

$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1-\lambda)\frac{\sum_{d,w} n(d,w)\log(\sum_k \theta_{dk}\beta_{kw})}{N_F} + \lambda\frac{\sum_{(u,d):v_{ud}\neq 0}(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1))}{N_V}$$
$$\boxed{-\frac{1}{2\sigma^2}(\sum_u ||\boldsymbol{x_u}||_2^2 + \sum_d ||\boldsymbol{a}_d||_2^2)}$$

Learning Algorithm

- An iterative algorithm where ideal points related parameters (X, A, b) and topic model related parameters (θ, β) enhance each other.
  - Step 1: Update X, A, b given θ, β
    - Gradient descent
  - Step 2: Update θ, β given X, A, b
    - Follow the idea of expectation-maximization (EM) algorithm: maximize a lower bound of the objective function in each iteration

$$
\begin{aligned}
&\sum_{d,w} n(d,w) \log \left(\sum_k \theta_{dk}\beta_{kw} \right) \\
=&\sum_{d,w} n(d,w) \log \left(\sum_k p(k|d,w) \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \right) \\
\geq&\sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \\
=&\sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \theta_{dk}\beta_{kw} - c
\end{aligned}
$$

Learning Algorithm

- Update θ: A nonlinear constrained optimization problem.
  Remove the constraints by a logistic function based transformation:

$$\theta_{dk} = \begin{cases} \dfrac{e^{\mu_{dk}}}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } 1 \le k \le K-1 \\ \dfrac{1}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } k = K \end{cases}$$

  and update μ_{dk} using gradient descent.
- Update β:
  Since β only appears in the topic model part, we use the same updating rule as in PLSA:

$$\beta_{kw}^{new} = \frac{\sum_d n(d,w) p(k|d,w)}{\sum_{d,w} n(d,w) p(k|d,w)} \qquad \text{where} \qquad p(k|d,w) = \frac{\theta_{dk} \beta_{kw}^{old}}{\sum_{k'} \theta_{dk'} \beta_{k'w}^{old}}$$

Outline

- Background and Problem Definition
- Topic Factorized Ideal Point Model and Learning Algorithm
- **Experimental Results**
- Conclusion

Data Description

- Dataset:
  - U.S. House and Senate roll call data in the years between 1990 and 2013.*
    - 1,540 legislators
    - 7,162 bills
    - 2,780,453 votes (80% are “YEA”)
  - Keep the latest version of a bill if there are multiple versions.
  - Randomly select 90% of the votes as training and 10% as testing.

* Downloaded from http://thomas.loc.gov/home/rollcallvotes.html

Evaluation Measures

- *Root mean square error (RMSE)* between the predicted vote score and the ground truth

$$\text{RMSE} = \sqrt{\frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)^2}{N_V}}$$

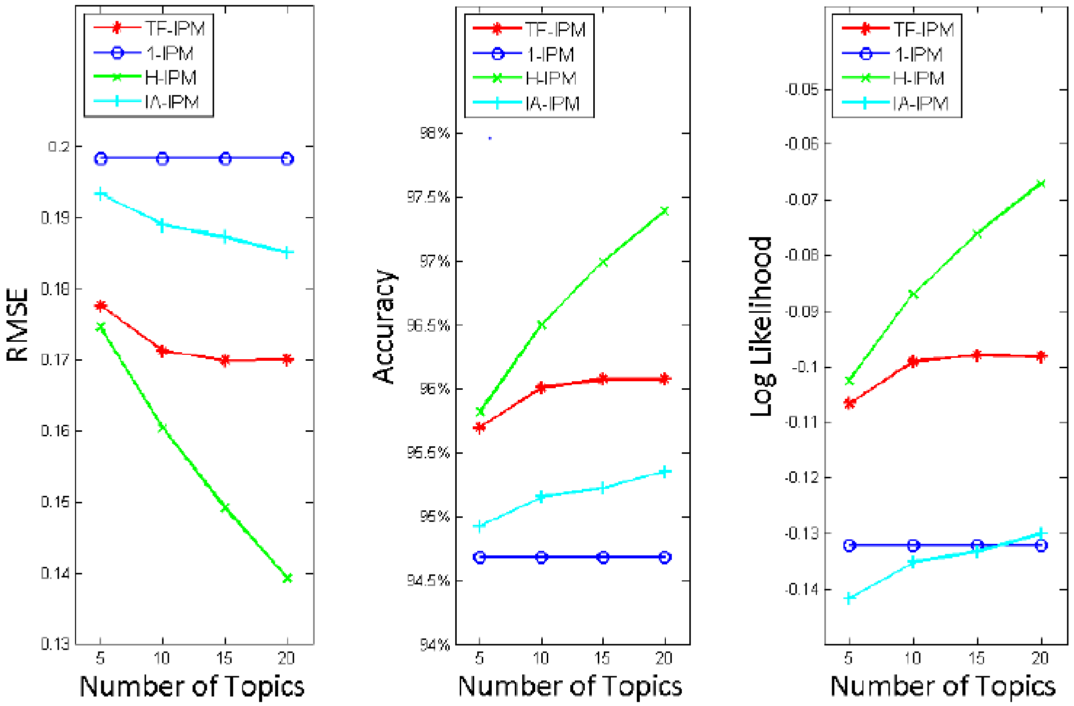
- *Accuracy* of correctly predicted votes (using 0.5 as a threshold for the predicted accuracy)

$$\text{Accuracy} = \frac{\sum_{u,d}\left(I_{\{p(v_{ud}=1)>0.5\ \&\&\ v_{ud}=1\}} + I_{\{p(v_{ud}=1)<0.5\ \&\&\ v_{ud}=-1\}}\right)}{N_V}$$

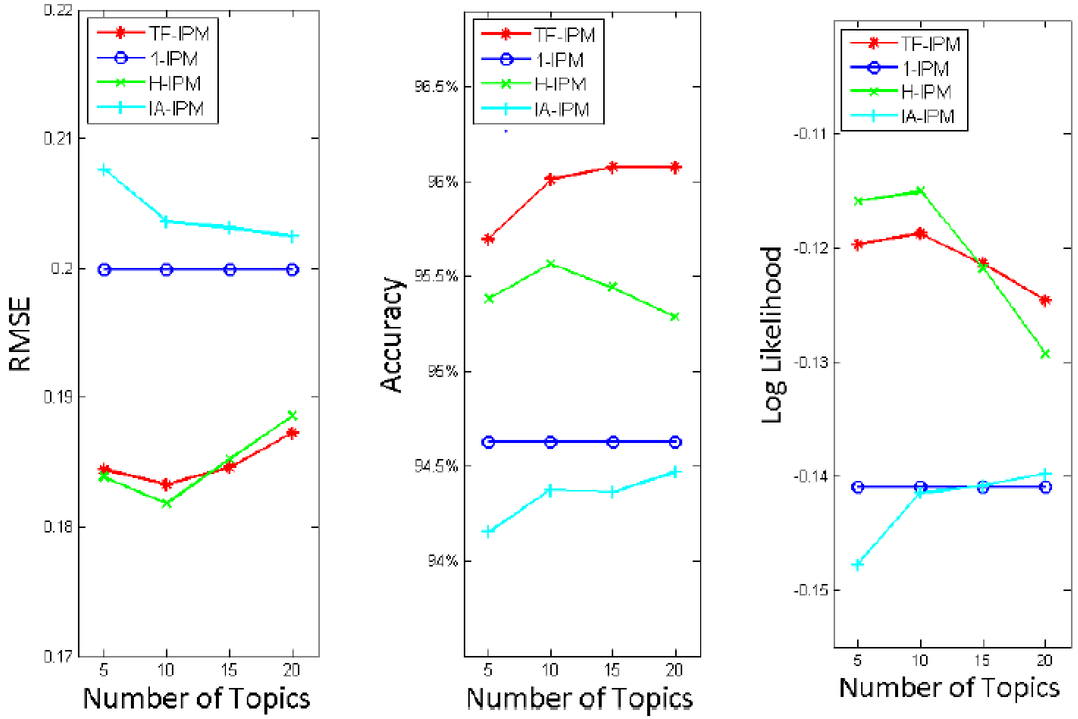
- *Average log-likelihood* of the voting link

$$\text{AvelogL} = \frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1)\right)}{N_V}$$

Experimental Results



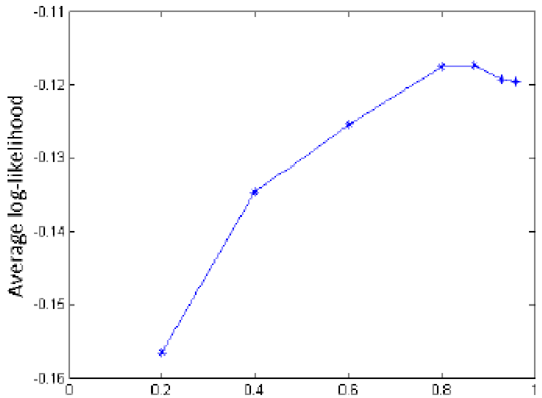
Training Data set



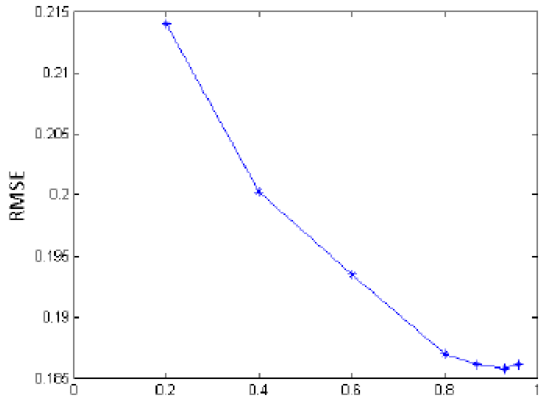
Testing Data set

Parameter Study

$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1 - \lambda) \cdot avelogL(text) + \lambda \cdot avelogL(voting) - \frac{1}{2\sigma^2}\left(\sum_u ||\boldsymbol{x_u}||_2^2 + \sum_d ||\boldsymbol{a}_d||_2^2\right)$$

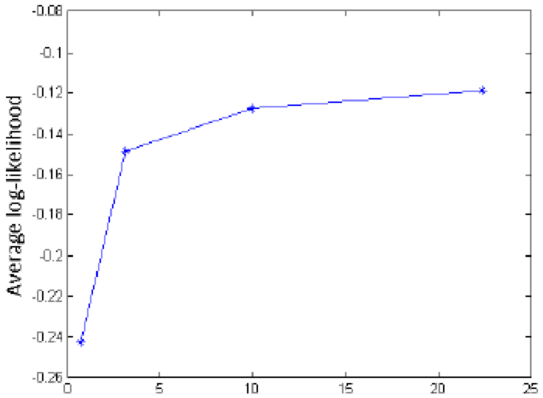


(a) Average log-likelihood

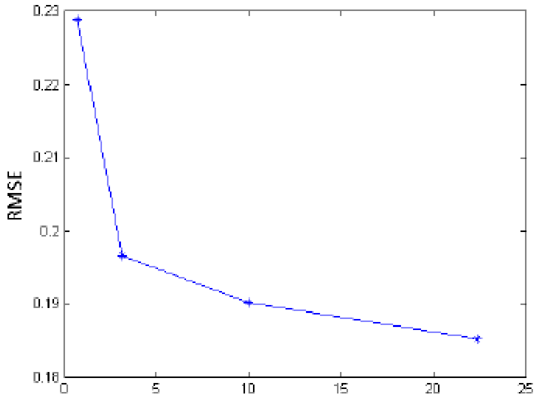


(b) RMSE

Parameter study on λ



(a) Average log-likelihood

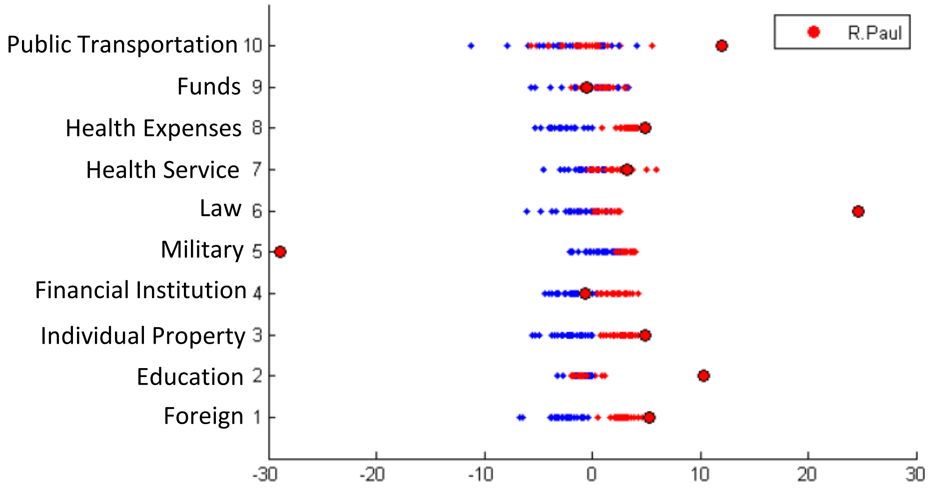


(b) RMSE

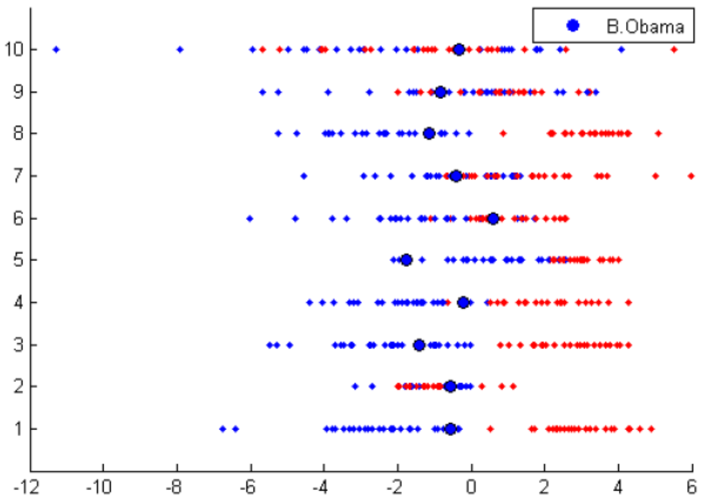
Parameter study on σ (regularization coefficient)

Case Studies

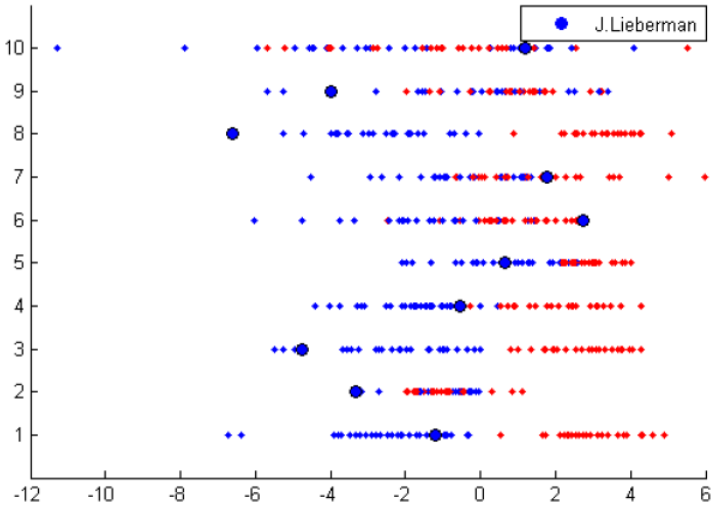
- Ideal points for three famous politicians: (Republican, Democrat)
  - Ronald Paul (R), Barack Obama (D), Joe Lieberman (D)



Ronald Paul



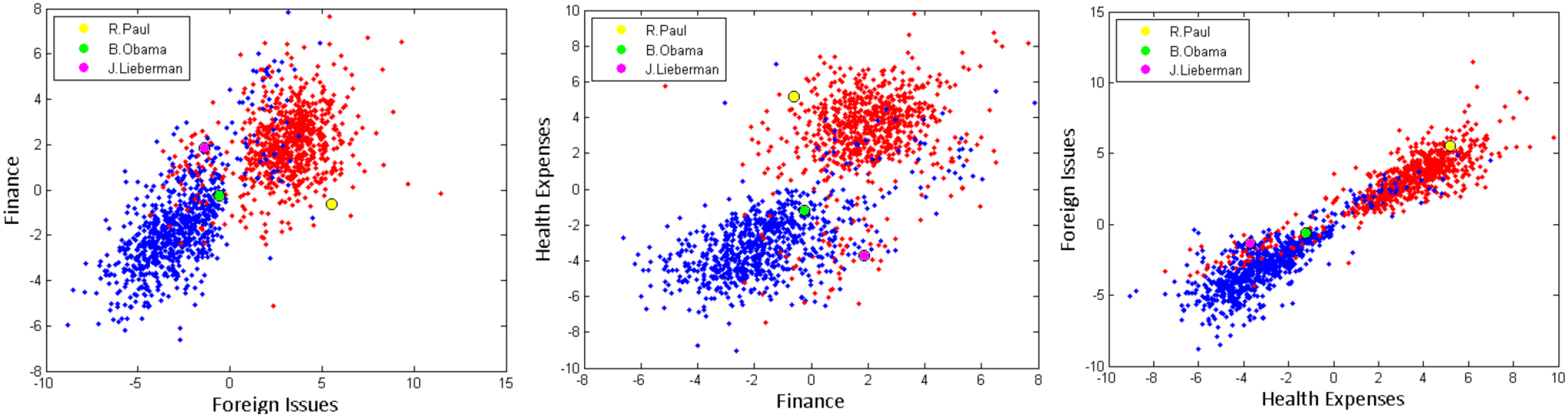
Barack Obama



Joe Lieberman

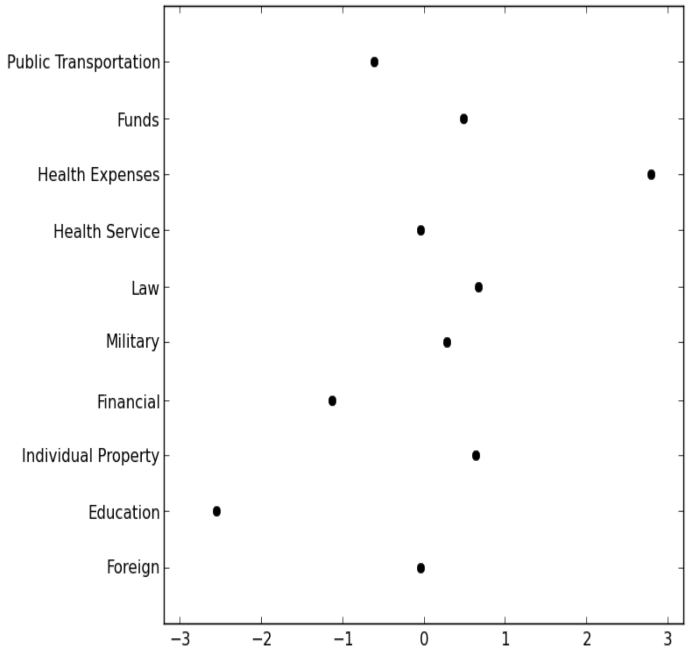
Case Studies

- Pick three topics: (Republican, Democrat)

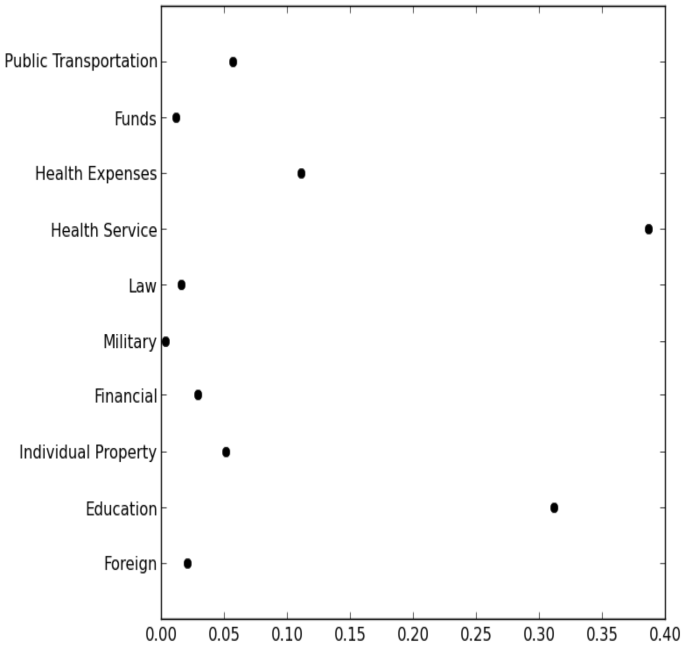


Case Studies

Bill: *H_R_4548* (in 2004)

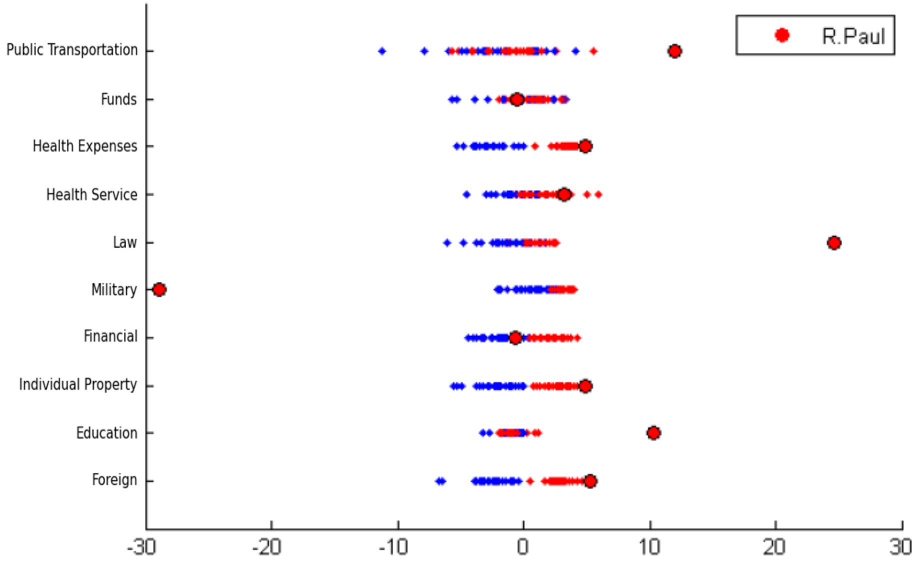


a_{dk}



θ_{dk}

$$p(v_{ud}=1)=\sigma(\sum_k \theta_{dk}x_{uk}a_{dk}+b_d)$$



x_{uk}

R. Paul $\xrightarrow{\text{Nay}}$ *H_R_4548*

For Unseen Bill d:

$\boxed{\text{Topic Model}} \rightarrow \boldsymbol{\theta}_d$

$\boxed{\text{TF-IPM}} \rightarrow \boldsymbol{x}_u$

$\boxed{\text{Experts}} \rightarrow \boldsymbol{a}_d, b_d$

$p(v_{ud}=1)$

Outline

- Background and Problem Definition
- Topic Factorized Ideal Point Model and Learning Algorithm
- Experimental Results
- **Conclusion**

Conclusion

- We estimate the ideal points of legislators and bills on multiple dimensions instead of global ones.
- The generation of topics are guided by two types of links in the heterogeneous network.
- We present a unified model that combines voting behavior and topic modeling, and propose an iterative learning algorithm to learn the parameters.
- The experimental results on real-world roll call data show our advantage over the state-of-the-art methods.

Thanks

Q & A