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

Ronald Paul

Barack Obama

Bill 1  Bill 2

......

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:
\(x_u\): Ideal Points for Politician \(u\)
\(a_d\): Ideal Points for Bill \(d\)

5

\(x_u\)'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.

\[
M \approx U \cdot V^T
\]

• 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

Entities:
• Politicians
• Bills
• Terms

Links:
• \((P, B)\)
• \((B, T)\)

Heterogeneous Voting Network
Text Part

Politicians

Bills

Terms
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)

\[
\theta_{dk} = p(k|d) \\
\beta_{kw} = p(w|k)
\]

\[
w_d = (n(d, 1), n(d, 2), ..., n(d, N_w))
\]

\[
p(w_d | \theta, \beta) = \prod_w \left( \sum_k \theta_{dk} \beta_{kw} \right)^{n(d,w)}
\]

\[
p(W | \theta, \beta) = \prod_d \prod_w \left( \sum_k \theta_{dk} \beta_{kw} \right)^{n(d,w)}
\]
Voting Part

Politicians → Bills → Terms
Voting Part

User-Bill voting matrix $V$

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_{ND}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$u_2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u_{NU}$</td>
<td>1</td>
<td>1 1 1</td>
</tr>
</tbody>
</table>

Voter $u$ $x_u$

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>......</th>
<th>Topic $k$</th>
<th>......</th>
<th>Topic $K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{u1}$</td>
<td>$x_{u2}$</td>
<td>......</td>
<td>$x_{uk}$</td>
<td>......</td>
<td>$x_{uK}$</td>
</tr>
</tbody>
</table>

Bill $d$ $a_d$

| $a_{d1}$ | $a_{d2}$ | ...... | $a_{dk}$ | ...... | $a_{dK}$ |

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

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

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

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

$$p(V|\theta, X, A, b) = \prod_{(u,d):v_{ud} \neq 0} \left( p(v_{ud} = 1) \frac{1 + v_{ud}}{2} p(v_{ud} = -1) \frac{1 - v_{ud}}{2} \right)$$
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.

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

s.t.

\[
0 \leq \theta_{dk} \leq 1, \quad \sum_k \theta_{dk} = 1 \quad \text{and} \quad 0 \leq \beta_{kw} \leq 1, \quad \sum_w \beta_{kw} = 1
\]

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

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

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

\[
\sum_{d,w} n(d, w) \log \left( \sum_k \theta_{dk} \beta_{kw} \right)
= \sum_{d,w} n(d, w) \log \left( \sum_k p(k|d, w) \frac{\theta_{dk} \beta_{kw}}{p(k|d, w)} \right)
\geq \sum_{d,w} n(d, w) \sum_k p(k|d, w) \log \frac{\theta_{dk} \beta_{kw}}{p(k|d, w)}
= \sum_{d,w} n(d, w) \sum_k p(k|d, w) \log \theta_{dk} \beta_{kw} - c
\]
Learning Algorithm

• Update $\theta$: A nonlinear constrained optimization problem.
  Remove the constraints by a logistic function based transformation:

\[
\theta_{dk} = \begin{cases} 
\frac{e^{\mu_{dk}}}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } 1 \leq k \leq K - 1 \\
\frac{1}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } k = K 
\end{cases}
\]

and update $\mu_{dk}$ using gradient descent.

• Update $\beta$:
  Since $\beta$ only appears in the topic model part, we use the same updating rule as in PLSA:

\[
\beta_{kw}^{new} = \frac{\sum_d n(d, w)p(k|d, w)}{\sum_{d, w} n(d, w)p(k|d, w)} \quad \text{where} \quad p(k|d, w) = \frac{\theta_{dk}\beta_{kw}^{old}}{\sum_{k'} \theta_{dk'}\beta_{k'w}^{old}}
\]
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)
  
  \[
  \text{Accuracy} = \frac{\sum_{u,d} \{ I[p(v_{ud} = 1) > 0.5 \&\& v_{ud} = 1] + I[p(v_{ud} = 1) < 0.5 \&\& v_{ud} = -1] \} }{N_V}
  \]

- **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

\[ J(\theta, \beta, X, A, b) = (1 - \lambda) \cdot \text{avelogL(text)} + \lambda \cdot \text{avelogL(voting)} - \frac{1}{2\sigma^2} \left( \sum_u ||x_u||^2_2 + \sum_d ||a_d||^2_2 \right) \]
Case Studies

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

• Pick three topics: (Republican, Democrat)
Case Studies

Bill: H_R_4548 (in 2004)

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

For Unseen Bill \(d\):

- Topic Model \(\theta_d\)
- TF-IPM \(x_u\)
- Experts \(a_{d}, b_d\)

\[ p(v_{ud} = 1) \]
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