Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network*

Yupeng Gu
Northeastern University
Boston, MA
ypgu@ccs.neu.edu

Yizhou Sun
Northeastern University
Boston, MA
yzsun@ccs.neu.edu

Ning Jiang
University of Illinois at Urbana-Champaign
Urbana, IL
jiang20@illinois.edu

Bingyu Wang
Northeastern University
Boston, MA
wang.bin@husky.neu.edu

Ting Chen
Northeastern University
Boston, MA
tingchen@ccs.neu.edu

ABSTRACT

Ideal point estimation that estimates legislators' ideological positions and understands their voting behavior has attracted studies from political science and computer science. Typically, a legislator is assigned a global ideal point based on her voting or other social behavior. However, it is quite normal that people may have different positions on different policy dimensions. For example, some people may be more liberal on economic issues while more conservative on cultural issues.

In this paper, we propose a novel topic-factorized ideal point estimation model for a legislative voting network in a unified framework. First, we model the ideal points of legislators and bills for each topic instead of assigning them to a global one. Second, the generation of topics are guided by the voting matrix in addition to the text information contained in bills. A unified model that combines voting behavior modeling and topic modeling is presented, and an iterative learning algorithm is proposed to learn the topics of bills as well as the topic-factorized ideal points of legislators and bills. By comparing with the state-of-the-art ideal point estimation models, our method has a much better explanation power in terms of held-out log-likelihood and other measures. Besides, case studies show that the topic-factorized ideal points coincide with human intuition. Finally, we illustrate how to use these topic-factorized ideal points to predict voting results for unseen bills.

Keywords

Ideal point estimation, legislative voting network, topic model, voting prediction

*The work was supported in part by the Yahoo ACE (Academic Career Enhancement) Award funding. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

KDD'14, August 24–27, 2014, New York, NY, USA.

http://dx.doi.org/10.1145/2623330.2623700 .

1. INTRODUCTION

Estimating voters' ideological positions is an important task, which can help us understand and predict their voting or other social behavior. For example, once we know a voter is more conservative, we can expect she is more likely to vote for republican candidates or vote against the bills that are very liberal. Ideal point estimation that estimates legislators' ideological positions has attracted studies from political science [4, 5, 8, 19] and computer science [6, 7], thanks to the widely available online voting records and other social behavior embedded in networks.

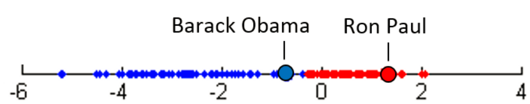


Figure 1: Illustration of standard one-dimensional ideal point model for sample Democrats (blue) and Republicans (red). B. Obama and R. Paul are highlighted.

In standard one-dimensional ideal point estimation models, a legislator is usually assigned a global ideal point based on her legislative voting records [4]. For example, in Fig. 1 we can see the ideal points of B. Obama and R. Paul according to their voting records when they were senator and house representative, respectively. From the figure, it is quite clear that Obama is liberal and Paul is conservative. The figure can also help to explain why Obama garnered support from some moderate republicans and conservatives during the 2008 election, and why Paul is regarded as one of the most conservative republicans in congress.

However, one-dimensional ideal point could be rather coarse to capture the whole picture of ideology, as it is quite normal that people may have different positions on different policy dimensions. For example, some people may be more liberal on economic issues while more conservative on cultural issues. From Fig. 2 we can see that the ideal points of Obama and Paul are rather different across a wide range of topics. In particular, the figure confirms the notion that Paul is the most conservative on the role of government on economy but not on social issues and foreign policy matters, especially as an opponent of war. Our goal of this paper is to estimate the ideal points of US House representatives and Senators in different topics, based on the roll call voting data from the 101st congress until now.

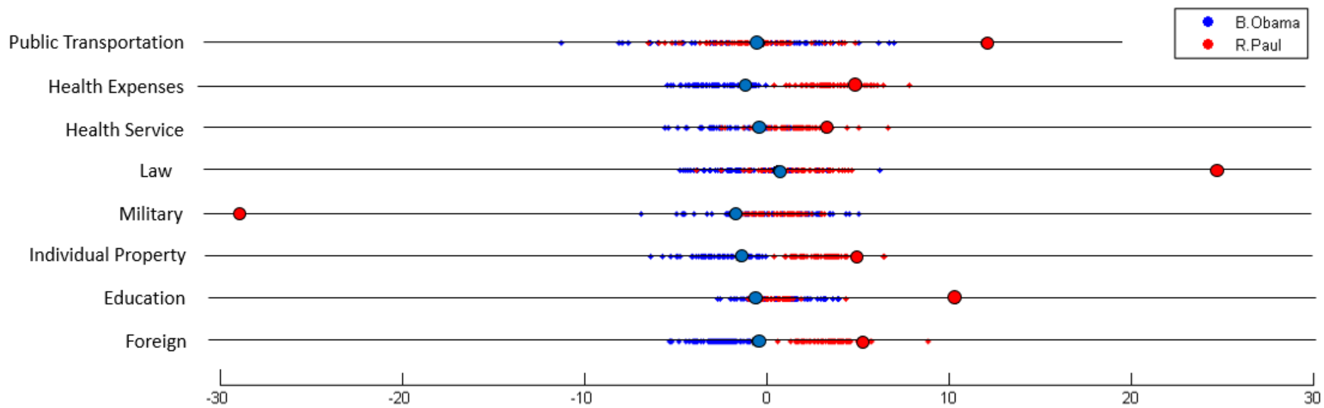


Figure 2: Illustration of topic-factorized ideal point model for sample Democrats (blue) and Republicans (red). B. Obama and R. Paul are highlighted.

A recent study [7] has shown that by incorporating the separately predetermined topic information of bills and modeling issue-adjusted ideal point for voters, the model has a much higher explanation power of the voting behavior. However, in this study, the topics are learned in advance, topic modeling and voting behavior modeling are two separate steps; and there is only a relative ideal point score in each topic. In this paper, we propose a novel topic-factorized ideal point estimation model by utilizing the legislative voting network data that contains (1) voting records between voters and bills, and (2) text information of the bills. The key idea is, on one hand, using topic and text information from the topic model can help enhance the voting link modeling; on the other hand, using voting data can help us gain a better understanding about the latent topics, therefore build a better latent topic distribution over text. Our approach enhances the issue-adjusted ideal point model in two-folds. First, we model the ideal point of voters and bills for each topic such that they can be comparable in different topic dimensions. Second, the generation of topics are guided by the voting matrix in addition to the text information in bills. A unified probabilistic model that combines voting behavior modeling and topic modeling is proposed. An iterative learning algorithm is proposed to learn the topics as well as the topic-factorized ideal points alternatively.

By comparing with the state-of-the-art ideal point models, our method has a much better explanation power in terms of held-out predictive accuracy. Besides, case studies show that the topic-factorized ideal points coincide with human intuition. Finally, we illustrate how to use these topic-factorized ideal points to make voting predictions for unseen bills.

The main contributions of our method can be summarized below.

- We propose a novel topic-factorized ideal point model based on text-rich legislative voting network, which simultaneously identifies the topics as well as the ideal points of voters and bills in a wide range of topics.

- An efficient algorithm is proposed to learn the model, which alternatively updates the topic distribution of each bill and ideal points for each voter and bill.

- The experiments are preformed on real-world US legislative roll call data. The results show that our model is superior to the state-of-the-art methods in several aspects.

The rest of this paper is organized as follows. In Section 2 we introduce the preliminaries and define the problem. In Section 3 we discuss our model in detail and introduce the learning algorithm. In Section 4 we show the experimental results. We summarize the related work in Section 5 and draw a conclusion of the paper in Section 6.

2. PRELIMINARY AND PROBLEM DEFINITION

In this section, we introduce the legislative voting network in our problem setting, the preliminary related to ideal point model (IPM), and define our problem formally.

2.1 Legislative Voting Network

A text-rich legislative voting network can be extracted from roll call voting data, which is comprised of three different types of objects: voters, bills, and words. Voters (u) and bills(d) are linked together according to the voting behavior, and the weight of links (v_{ud}) is either 1 (“YEA”) or -1 (“NAY”). Bills and words are linked together if the word appears in the bill, and the weight is determined by the number of occurrences of the word. An example of such legislative voting network that is extracted from US roll call voting data is shown in Fig. 3. It is worth noting that the value of v_{ud} between user u and bill d can also be 0, which means the link between u and d is missing. We do not model missing links in this paper.

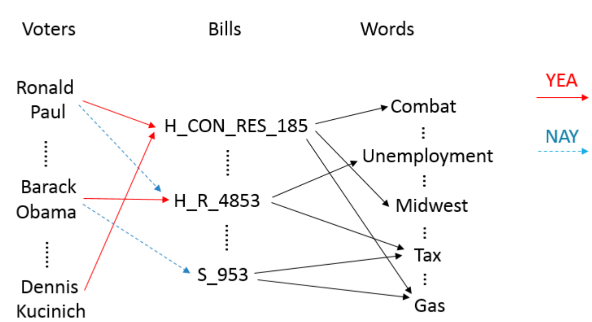


Figure 3: An illustration of legislative voting network.

2.2 Ideal Point Models

The goal of ideal point model is to estimate the policy positions of voters and bills that can explain the voting or other social behavior. Poole and Rosenthal [17] proposed a seminal one-dimensional ideal point model for understanding legislative behavior. In [19], Poole and Rosenthal studied a similar model in a higher-dimensional setting. Alternative higher-dimensional ideal point models have been developed by Clinton et al. [4], Heckman and Snyder [8], and Londregan [12]. We first briefly introduce one-dimensional and higher-dimensional ideal point models, then the closely-related issue-adjusted ideal point model by Blei and Gerrish [7] that utilized text information in bills to make topics as explicit dimensions.

One-dimensional ideal point model.

In one-dimensional ideal point model, each legislator u is associated with an ideal point x_u, which represents her position. Each bill d has a polarity a_d and a popularity offset b_d as well. The probability of voting "YEA" is given by $p(v_{ud} = 1) = \sigma(x_u \cdot a_d + b_d)$, where σ denotes the logistic or probit function. All the parameters are estimated by maximizing the likelihood of observing the voting matrix under this model.

Higher-dimensional ideal point model.

Due to the limitation that one-dimensional ideal point cannot capture the different behavior of legislators for different topics, some researches extend the above model to higher dimensions. For example, Clinton et al. [4] proposed to map each voter and bill into a K-dimensional space, and model the probability of a voter voting "YEA" for a bill with the probability $\sigma(\mathbf{x}_u \cdot \mathbf{a}_d + b_d)$, where $\mathbf{x}_u$ and $\mathbf{a}_d$ are in the space of $\mathbb{R}^K$ and " $\cdot$ " denotes dot product.

Issue-adjusted ideal point model.

The latent space derived in the higher-dimensional ideal point model, however, is very difficult to interpret. Thus, an issue-adjusted ideal point model [7] is proposed very recently to estimate the adjusted ideal points in some separately predetermined issues (topics). In this model, every legislator u has an adjusted issue preference vector $\mathbf{z}_u \in \mathbb{R}^K$ (K is the number of topics) beyond the global ideal point x_u. The probability of voting "YEA" is determined by her global ideal point as well as the adjusted ideal point in each topic, weighted by the conditional expectation of a bill's topic distribution: $p(v_{ud} = 1) = \sigma((\mathbf{z}_u^T \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{w}_d] + x_u)a_d + b_d)$, where $\boldsymbol{\theta}_d$ is the topic distribution of a bill d and $\mathbf{w}_d$ is the observed word vector for bill d.

In this model, however, the topics are predetermined, and there is no impact from voting records which in fact can effectively enhance the topic generation. Besides, since ideal points for all topics are adjusted based on the same global point, the model puts constraints on how a voter can change her positions in different topics. We thus propose a new ideal point model that can overcome the two limitations, and the problem is defined formally in the following.

2.3 Problem Definition

Given a legislative voting network G, which contains N_U voters, N_D bills and N_W terms, and the number K of topics, our goal is to learn the topic models ($\boldsymbol{\Theta} = \{\boldsymbol{\theta}_d\}_{d=1}^{N_D}$, $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}_{k=1}^{K}$) and estimate topic-factorized ideal points for voters ($\mathbf{X} = \{x_{uk}\}$) and bills ($\mathbf{A} = \{a_{dk}\}$), which can best explain the observed network. Here, we use x_{uk} to denote the ideal point of voter u in topic k, and a_{dk} to denote the ideal point of bill d in topic k.

The detailed topic-factorized ideal point model that can solve this problem is introduced in next section, and all notations used can be found in Table 1.

Table 1: Table of Notation

N_U	Total number of voters
N_D	Total number of bills
N_W	Total number of terms
N_V	Total number of votes
K	Total number of topics
$\boldsymbol{\theta}_d$	Topic distribution for bill d
$\boldsymbol{\beta}_k$	Word distribution for topic k
x_{uk}	Ideal point for voter u in topic k
a_{dk}	Ideal point for bill d in topic k
$n(d, w)$	Frequency of word w appearing in bill d
v_{ud}	Voting from voter u to bill d
b_d	Constant offset of bill d
$\mathbf{V}$	The observed voting matrix $\{v_{ud}\}$
$\mathbf{W}$	The observed document-word matrix $\{n(d, w)\}$
$\boldsymbol{\Theta}$	The parameter matrix $\{\theta_{dk}\}$
$\mathbf{X}$	The parameter matrix $\{x_{uk}\}$
$\mathbf{A}$	The parameter matrix $\{a_{dk}\}$

3. OUR APPROACH

In this section, we first introduce the topic-factorized ideal point model (TF-IPM) in detail, and then propose the learning algorithm that can learn the parameters in the model given the observed legislative voting network.

3.1 The Topic-Factorized Ideal Point Model

The idea of our model is to build a probabilistic model for the legislative voting network, which contains two parts: (1) the probability modeling for observing the whole corpus of bills given the topic models, and (2) the probability modeling for the voting record given the topic distribution of bills as well as the ideal points for bills and voters in each topic. The goal is then to learn the parameters including the topic distributions and ideal points that can maximize the combined likelihood function.

The topic-factorized ideal point model is illustrated by a graphical model as shown in Figure 4. We now introduce the two parts separately in the following.

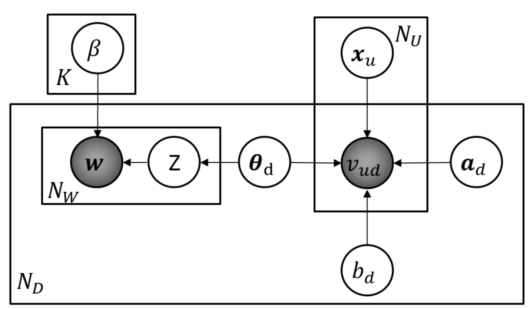


Figure 4: Graphical model for topic-factorized ideal point model.

Modeling the text part.

Following the idea of traditional topic models PLSA [10] and LDA [3], we can model the probability of each word in each document as a mixture of multinomial distributions. Let $n(d, w)$ be the frequency of term w appearing in bill d, $\theta_{dk} = p(k|d)$ be the probability of document d belonging to topic k, and $\beta_{kw} = p(w|k)$ be the probability that topic k generates word w, then the probability of word w appearing in document d with $n(d, w)$ occurrences can be defined as $(\sum_k \theta_{dk}\beta_{kw})^{n(d,w)}$. The probability of observing the word count vector $\mathbf{w}_d = (n(d,1), n(d,2), \ldots, n(d, N_W))$ for document d is then $p(\mathbf{w}_d|\mathbf{\Theta}, \boldsymbol{\beta}) = \prod_w (\sum_k \theta_{dk}\beta_{kw})^{n(d,w)}$, and the probability of observing the whole corpus is then:

$$p(\mathbf{W}|\mathbf{\Theta}, \boldsymbol{\beta}) = \prod_d \prod_w (\sum_k \theta_{dk}\beta_{kw})^{n(d,w)} \qquad (1)$$

where $\mathbf{W}$ is the observed document-word count matrix.

Modeling voting links.

Now we introduce how the voting links can be modeled. Intuitively, legislator (u) are associated with different ideal points in different topics, which are represented using a K-dimensional vector $\mathbf{x}_u$; and bills (d) are also associated with different ideal points if they belong to different topics, which are represented using another K-dimensional vector $\mathbf{a}_d$. For each vote issued from a user u to a bill d, the probability of whether the $vote$ is an "YEA" or "NAY" is determined by the ideal points for u and d in each topic. For example, if the bill is a combination of "business" and "healthcare", we need to see whether the ideal points of u and d agree with each other in both topics.

Let v_{ud} be the vote score from voter u to bill d, we use $v_{ud} = 1$ to denote "YEA", i.e., u voted for d, and use $v_{ud} = -1$ to denote "NAY", i.e., u voted against d. The probabilities of the two values of v_{ud} are defined by:

$$\begin{aligned} p(v_{ud} = 1) &= \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d) \\ p(v_{ud} = -1) &= 1 - \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d) \end{aligned} \qquad (2)$$

where θ_{dk} is the topic proportion for bill d in topic k, x_{uk} denotes the ideal point of voter u in topic k, a_{dk} denotes the ideal point of bill d in topic k, b_d denotes the constant offset, and σ denotes the logistic function. Since $\sum_k \theta_{dk} = 1$, the summation $\sum_k \theta_{dk} x_{uk} a_{dk}$ can be regarded as the weighted average of the agreement between legislator u and bill d in K topics. Note that, we only model the observed links from voters to bills, i.e., for the pairs (u, v) that $v_{ud} \neq 0$. In other words, whether a voter u votes a bill d is not modeled, as in reality legislators cast their votes for almost all bills in her term.

The overall likelihood of observed votes can be modeled as:

$$p(\mathbf{V}|\mathbf{\Theta}, \mathbf{X}, \mathbf{A}, \mathbf{b}) = \prod_{(u,d):v_{ud}\neq 0} \left(p(v_{ud} = 1)^{\frac{1+v_{ud}}{2}} p(v_{ud} = -1)^{\frac{1-v_{ud}}{2}} \right) \qquad (3)$$

where $\mathbf{V}$ is observed voting matrix, $\mathbf{\Theta}, \mathbf{X}, \mathbf{A}, \mathbf{b}$ are parameters in matrix/vector form for topic distribution, ideal points, and constant factors. $\frac{1+v_{ud}}{2}$ and $\frac{1-v_{ud}}{2}$ transforms the original votes to 1 or 0, which plays a role as an indicator function.

Putting it together.

We now combine the two parts together, and the final objective function is a linear combination of the two average log likelihood functions over the word links and voting links. In addition, we also add an l_2 regularization term to $\mathbf{A}$ and $\mathbf{X}$ to reduce the effect of over-fitting, equivalently by assuming they are from a Gaussian prior with mean as 0, and standard deviation as σ:

$$\begin{aligned} J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \mathbf{b}) = (1-\lambda) \frac{\sum_{d,w} n(d,w) \log \left(\sum_k \theta_{dk}\beta_{kw}\right)}{\sum_{d,w} n(d,w)} + \\ \lambda \frac{\sum_{(u,d):v_{ud}\neq 0} \left(\frac{1+v_{ud}}{2} \log p(v_{ud}=1) + \frac{1-v_{ud}}{2} \log p(v_{ud}=-1)\right)}{N_V} \\ - \frac{1}{2\sigma^2} \cdot (\sum_{u,k} x_{uk}^2 + \sum_{d,k} a_{dk}^2) \end{aligned} \qquad (4)$$

s.t.

$$0 \leq \theta_{dk} \leq 1, \qquad \sum_k \theta_{dk} = 1$$

and

$$0 \leq \beta_{kw} \leq 1, \qquad \sum_w \beta_{kw} = 1$$

where $\sum_{d,w} n(d,w)$ denotes the total number of word occurrences in the bills, N_V denotes the total number of votes, and $\lambda \in (0, 1)$ is the tradeoff weight over the two average log-likelihood functions. The goal is now to find the best parameters $\mathbf{\Theta}, \boldsymbol{\beta}, \mathbf{X}, \mathbf{A}, \mathbf{b}$ that maximize the objective function.

Discussions.

The Advantage of Modeling $\mathbf{\Theta}$

In issue-adjusted IPM, a pre-calculated $\mathbf{\Theta}$ has been used as an auxiliary parameter in the voting part. In practice, however, voting behaviors can also help to guide the topics generation. From our experiments, we can see that the held-out predictive accuracy can be significantly enhanced when topics are further adjusted according to voting behaviors.

Topic Model with Background

In [23], a background model is proposed to help improve the PLSA model. This background model can detect topics that are more distinctive. As there are many background words in the bills, we find that by including a background model in the topic modeling part, the performance of our model can be further enhanced. Considering to compare with issue-adjusted model more fairly, we do not introduce the results in the experiment section.

Adding Priors to TF-IPM

In issue-adjusted IPM, priors are added on the parameters. For example, $\boldsymbol{X}$ is assumed from a Gaussian distribution. Similar priors are put on the parameters of our model via the regularization term. In the experiment section, in order to compare all the models fairly, we put the same regularization term for all the baselines.

3.2 The Learning Algorithm

In this section, we introduce the learning algorithm to estimate the parameters, which is an iterative algorithm containing two steps. In the first step, we update the ideal points related parameters ($\mathbf{X}$, $\mathbf{A}$, and $\mathbf{b}$) when topics related parameters ($\mathbf{\Theta}$ and $\boldsymbol{\beta}$) are fixed; and in the second step, we update the topic models ($\mathbf{\Theta}$ and $\boldsymbol{\beta}$) when fixing ideal points related parameters ($\mathbf{X}$, $\mathbf{A}$, and $\mathbf{b}$).

The first step: updating $\mathbf{X}$, $\mathbf{A}$, $\mathbf{b}$.

We estimate $\mathbf{X}$, $\mathbf{A}$, $\mathbf{b}$ using gradient descent algorithm, which involves derivative calculation in the objective function. They are updated using the following rule:

$$
\begin{aligned}
x_{uk} &= x_{uk} + \eta \frac{\partial J}{\partial x_{uk}} \\
a_{dk} &= a_{dk} + \eta \frac{\partial J}{\partial a_{dk}} \\
b_d &= b_d + \eta \frac{\partial J}{\partial b_d}
\end{aligned} \tag{5}
$$

where η is the learning rate (we initialize η to be 10^{-5}, and use line search strategy to adjust η automatically), and

$$
\begin{aligned}
\frac{\partial J}{\partial x_{uk}} &= \lambda \frac{\sum_{d:v_{ud}\neq 0} a_{dk} x_{uk} \left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)}{N_V} - \frac{x_{uk}}{\sigma^2} \\
\frac{\partial J}{\partial a_{dk}} &= \lambda \frac{\sum_{u:v_{ud}\neq 0} \theta_{dk} x_{uk} \left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)}{N_V} - \frac{a_{dk}}{\sigma^2} \\
\frac{\partial J}{\partial b_d} &= \lambda \frac{\sum_{u:v_{ud}\neq 0} \left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)}{N_V}
\end{aligned} \tag{6}
$$

The second step: updating $\mathbf{\Theta}$ and $\boldsymbol{\beta}$.

Since the summation of θ is inside log function, the derivative is not easy to compute. Here we follow the idea of expectation-maximization (EM) algorithm and maximize a lower bound of the original objective function. We have noted that

$$
\begin{aligned}
&\sum_{d,w} n(d,w) \log \left(\sum_k \theta_{dk} \beta_{kw} \right) \\
=& \sum_{d,w} n(d,w) \log \left(\sum_k p(k|d,w) \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \right) \\
\geq& \sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \\
=& \sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \theta_{dk}\beta_{kw} - c
\end{aligned} \tag{7}
$$

where $c = \sum_{d,w} n(d,w) \sum_k p(k|d,w) \log p(k|d,w)$ is a constant with respect to θ and β. The inequality holds because of the Jensen's inequality of the concave function $\log(x)$. We denote this lower bound as:

$$
\begin{aligned}
J_1(\mathbf{\Theta}, \boldsymbol{\beta}) =& (1-\lambda) \frac{\sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \theta_{dk}\beta_{kw}}{\sum_{d,w} n(d,w)} + \\
&\lambda \frac{\sum_{u,d} \left(\frac{1+v_{ud}}{2} \log p(v_{ud}=1) + \frac{1-v_{ud}}{2} \log p(v_{ud}=-1)\right)}{N_V} \\
&- \frac{1}{2\sigma^2} \cdot \left(\sum_k x_{uk}^2 + \sum_k a_{dk}^2\right)
\end{aligned} \tag{8}
$$

and now we seek to maximize it w.r.t $\mathbf{\Theta}$ and $\boldsymbol{\beta}$.

Updating $\mathbf{\Theta}$ given other parameters is a nonlinear constrained optimization problem. We could use the method of Lagrange multipliers to remove the constraints, but it involves to solve a complicated function of Lagrange multipliers. Here we use another method to remove the constraints. We regard θ_{dk} as a function of a new set of parameters μ_{dk} so that there is no constraints on $\boldsymbol{\mu}$. We define the logistic function-based transformation as

$$
\theta_{dk} = \begin{cases} \dfrac{e^{\mu_{dk}}}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & if\ 1 \leq k \leq K-1 \\[2ex] \dfrac{1}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & if\ k = K \end{cases} \tag{9}
$$

The new set of parameters $\mu_{dk} (1 \leq k \leq K-1)$ are real numbers and they have no constraints, therefore we can update μ_{dk} using gradient descent and recover θ_{dk} using the formula above.

The derivative w.r.t $\boldsymbol{\mu}$ is

$$
\frac{\partial J_1}{\partial \mu_{dk}} = \sum_{k'=1}^{K} \frac{\partial J_1}{\partial \theta_{dk'}} \frac{\partial \theta_{dk'}}{\partial \mu_{dk}} \tag{10}
$$

where

$$
\begin{aligned}
\frac{\partial J_1}{\partial \theta_{dk'}} =& \frac{1-\lambda}{\sum_{d,w} n(d,w)} \frac{\sum_w n(d,w) p(k'|d,w)}{\theta_{dk'}} \\
&+ \frac{\lambda}{N_V} \sum_u a_{dk'} x_{uk'} \left(\frac{1+v_{ud}}{2} - p(v_{ud}=1) \right)
\end{aligned} \tag{11}
$$

and

$$
\frac{\partial \theta_{dk'}}{\partial \mu_{dk}} = \begin{cases} \theta_{dk}(1-\theta_{dk}) & if\ k' = k \\ -\theta_{dk'}\theta_{dk} & if\ k' \neq k \end{cases} \tag{12}
$$

We observe that $\boldsymbol{\beta}$ only appears in the topic model part. So we update $\boldsymbol{\beta}$ using the EM algorithm as Hofmann proposed [10]. The update rule of $\boldsymbol{\beta}$ is

$$
\beta_{kw}^{new} = \frac{\sum_d n(d,w) p(k|d,w)}{\sum_{d,w} n(d,w) p(k|d,w)} \tag{13}
$$

where

$$
p(k|d,w) = \frac{\theta_{dk}^{old} \beta_{kw}^{old}}{\sum_{k'} \theta_{dk'}^{old} \beta_{k'w}^{old}} \tag{14}
$$

4. EXPERIMENTS

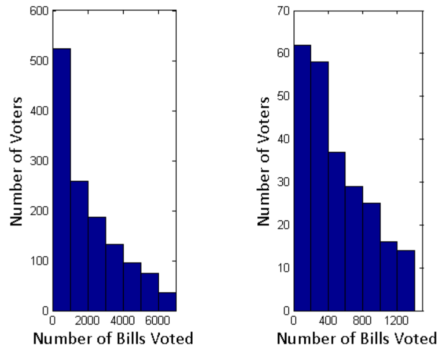
In this section, we show the experimental results of our model on the legislative voting network extracted from U.S. House and Senate roll call data.

4.1 Data Description

We collect the U.S. House and Senate roll call data in the years between 1990 and 2013 from THOMAS[1], a public website providing federal legislative voting information. Votes and text information for bills are downloaded. We only select the bills where the text information is available, and keep the latest version of a bill if there are multiple versions. For each bill, we first remove stop words, and then choose 10,000 distinct words with highest frequency as the dictionary. 1299 House representatives and 241 senators are collected. Legislators from House have voted 6479 bills, while legislators from Senate have voted 1247 bills. There are 564 bills that are voted by both House and Senate.

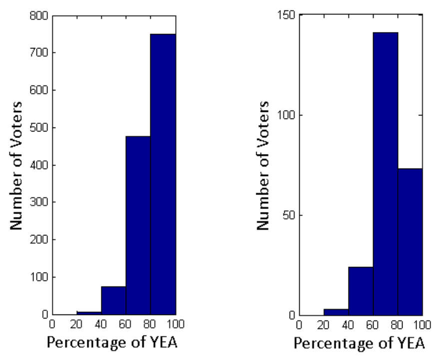
In total, there are 1540 legislators, 7162 bills, and 2780453 votes. Some interesting statistics of the data set are shown in Fig. 5 and 6.

[1] `http://thomas.loc.gov/home/rollcallvotes.html`



(a) House (b) Senate

Figure 5: Statistics of number of votes for legislators in House and Senate.



(a) House (b) Senate

Figure 6: Statistics of percentage of "YEA" for legislators in House and Senate.

4.2 Performance Evaluation

We now compare our TP-IPM model with several baseline methods in several aspects to demonstrate the power of the proposed model.

4.2.1 Baseline Methods

We compare our methods with three baseline methods.

1. One-dimensional ideal point model described in Section 2.2, denoted as 1-IPM.

2. High-dimensional ideal point model described in Section 2.2, denoted as H-IPM.

3. Issue-adjusted ideal point model described in Section 2.2, denoted as IA-IPM.

We use the same number of K for H-IPM, IA-IPM, and TF-IPM, to make sure that the dimensionality of idea points is the same.

4.2.2 Evaluation Measures

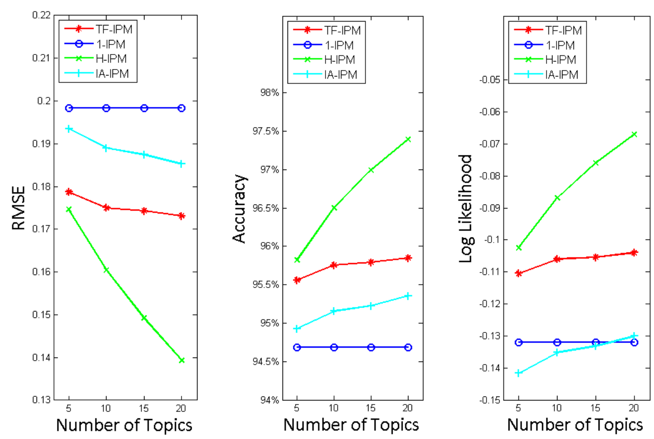
In order to see whether our model has good explanation power of the legislative network, we compare our model to others in terms of the following measures for both training and testing dataset, where we randomly select 90% of the votes as training and 10% of the votes as testing.

- Root-mean-square error (RMSE) between the predicted vote score and the ground truth vote score: $RMSE = \sqrt{\frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)^2}{N_V}}$.

- Accuracy of correctly predicted votes (using 0.5 as threshold for the predicted probability) among all the votes: $Accuracy = \frac{\sum(sgn(p(v_{ud}=1)-0.5)==v_{ud})}{N_V}$, where sgn is the sign function.

- Average log-likelihood of the voting link: $avglogL = \frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2}\log p(v_{ud}=1)+\frac{1-v_{ud}}{2}\log p(v_{ud}=-1)\right)}{N_V}$.

The total number of topics is set to be 10, and λ is set to be 0.8 in TF-IPM. The regularization parameter σ is chosen to be 22.4 in all methods so as to make the regularization coefficient $\frac{1}{2\sigma^2}$ around 0.001.



(a) RMSE (b) Accuracy (c) Average log-likelihood

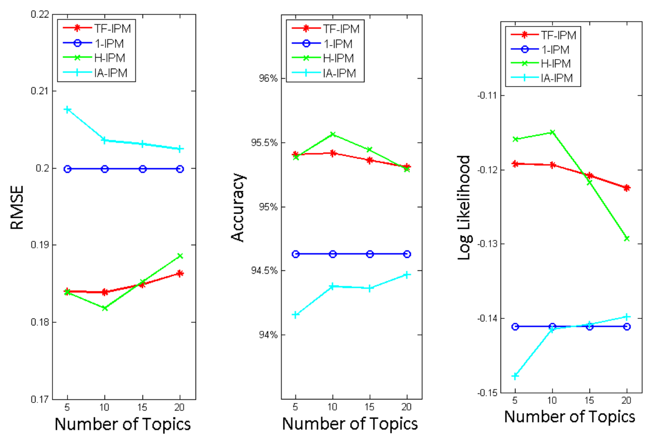
Figure 7: Comparison on training data set.

From the Figure 7, we can see that high-dimensional IPM (H-IPM) has the best explanation power over the training dataset in terms of all three measures, due to its flexibility in choosing any possible latent factors. TF-IPM has the second best training accuracy. However, in Figure 8, we can see that, TF-IPM can overcome the overfitting issue of H-IPM when number of topics increases and has the better testing performance in terms of all three measures than all the other two baselines. Considering the interpretability of topic-factorized ideal points, we can see TF-IPM is a better choice.

Most of the models suffer from over-fitting problems when the total number of topics increases. We keep using 10 topics in the remaining part of the paper.

4.2.3 Parameter Study

In our model we have used the parameter λ that links the likelihood of topic model part and voting part as well as the regularization parameter σ. Intuitively, a larger λ means we pose more emphasis on the voting part. We evaluate the effect of λ and show it in Fig. 9. By varying the value of λ in the range of $(0, 1)$, we can derive different parameter estimations, and calculate the average log-likelihood and RMSE on testing data. From the result, we can clearly see that voting part indeed plays a more important role than the text part, but when λ is too big, the performance of our

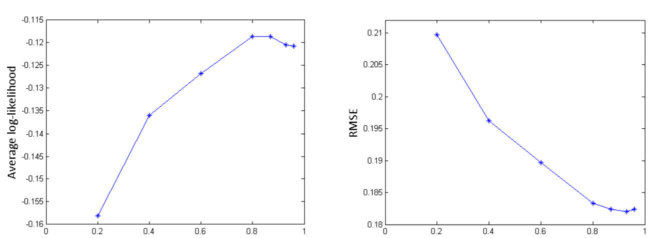


(a) RMSE (b) Accuracy (c) Average log-likelihood

Figure 8: Comparison on testing data set.

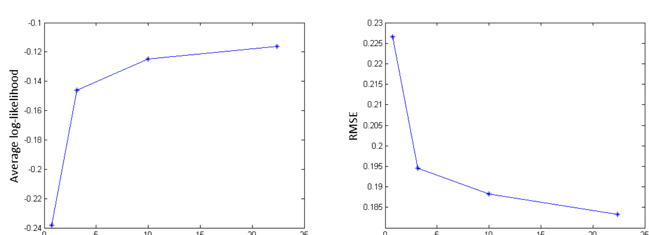
model will become worse due to overfitting. We set $\lambda = 0.8$ in all other experiments.

We also study the effect of the regularization term on the performance of our model. We try different values of σ and evaluate average log-likelihood and RMSE on testing dataset. The results are shown in Fig. 10, and we can see that the performance becomes less sensitive when σ becomes larger. We set $\sigma = 22.4$ in all other experiments.



(a) Average log-likelihood (b) RMSE

Figure 9: Parameter study on λ.



(a) Average log-likelihood (b) RMSE

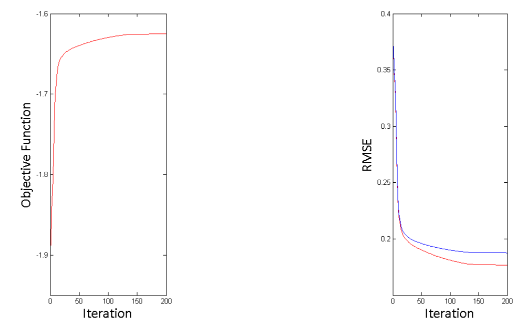
Figure 10: Parameter study on σ.

4.2.4 *Convergence of Our Learning Algorithm*

In order to get a better idea of our iterative learning algorithm, we study the convergence of the learning algorithm here. Fig. 11 shows that our learning algorithm converges. Along with the iterations, the objective function is increasing, while training and testing errors are decreasing.

Table 2: Selected top words in each topic

#	Post-label	Top words
1	Foreign	Foreign, Government, International
2	Education	Education, School, Students
3	Individual property	Individual, Property, Fiscal
4	Financial	Institution, Financial, Agency, Bank
5	Military	Military, Defense, Duty
6	Law	Law, Authority, Provision
7	Health service	Health, Service, Assistance
8	Health expenses	Health, Fiscal, Expenses
9	Funds	Funds, Fiscal, Amounts
10	Public transportation	Public, Transportation, Motor



(a) Objective Function (b) RMSE

Figure 11: Convergence Study

4.3 Case Studies

We now show some case studies to demonstrate ideal points detected by our model. By setting topic number $K = 10$ and $\lambda = 0.8$, we get topics as shown in Table 2.

We study the voting behaviors for three famous legislators: Ronald Paul, Republican House representative from Texas (in office: 1976-1977, 1979-1985, and 1997-2013); Barack Obama, Democratic Senator from Illinois (in office: 2005 - 2008); and Joe Lieberman, Democratic representative from Connecticut (in office: 1989 - 2013). Their ideal points are shown in Fig. 12, 13, and 14, respectively. We also show the ideal points for some other people in order to understand their relative position and make the scale more meaningful. We can see that: (1) Obama is in general in the middle and moderately left in almost all the topics; (2) Paul is regarded as one of the most conservative republicans in congress, and is most conservative on the role of government on economy but not on social issues and especially is an opponent of war; and (3) Lieberman is regarded as a Democrat who is conservative in social goods (e.g., transportation) and military issues.

We also select one bill that Ronald Paul has voted. It is a bill in 2006 and its number is H.RES.578. It mainly covers supporting the government of Romania to improve the standard health care and well-being of children in Romania. We can find the positions a_{dk} as well as the topic distribution θ_{dk} of this bill in Fig. 15. We can clearly see that the most notable topics the bill belongs to are “Health Service” and “Funds.” We can also infer from the ideal points of Ronald Paul and the ideal points for H.RES.578 that R.Paul is in favor of this bill.

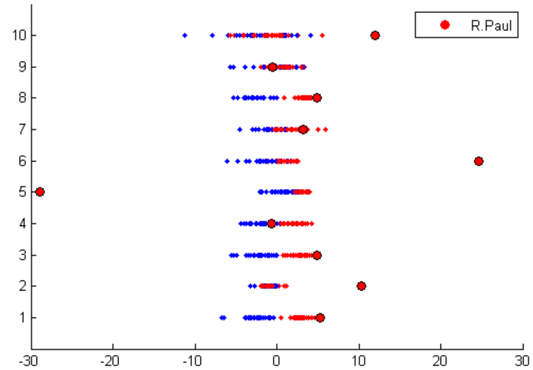


Figure 12: Estimated Ideal Points for Ronald Paul

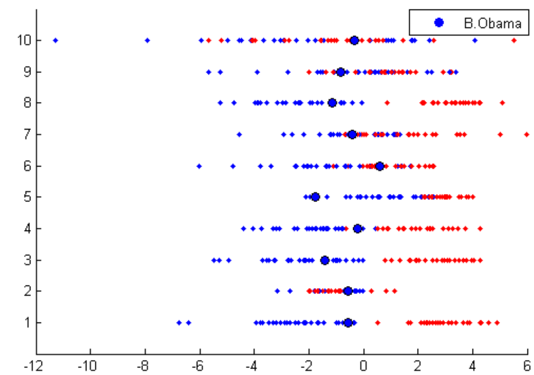


Figure 13: Estimated Ideal Points for Barack Obama

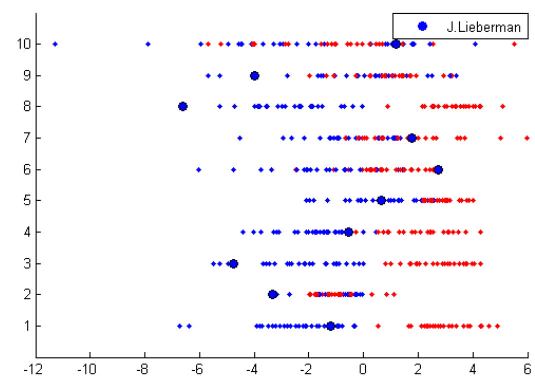
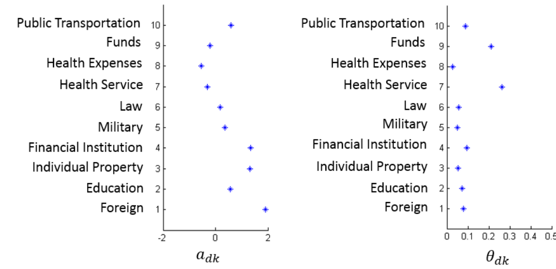


Figure 14: Estimated Ideal Points for Joe Lieberman

In addition, we select three topics, energy, education and individual properties, and draw scatter plots for all the legislators in every two of these three topics. Democratic people are represented as blue dots, while Republican people are represented as red ones. Especially, Ronald Paul is represented as a yellow circle, Barack Obama as a green circle, and Joe Lieberman as a pink circle. We can see clearly from Fig. 16, 17 and 18 that legislators in different parties lie on different sides of each plane. It can be seen that, by such analysis it is easy to figure out how close two legislators are in terms of certain topic dimensions. We can also easily detect outliers from such visualization.



(a) Bill Positions (b) Topic Distributions

Figure 15: Bill H_RES_578

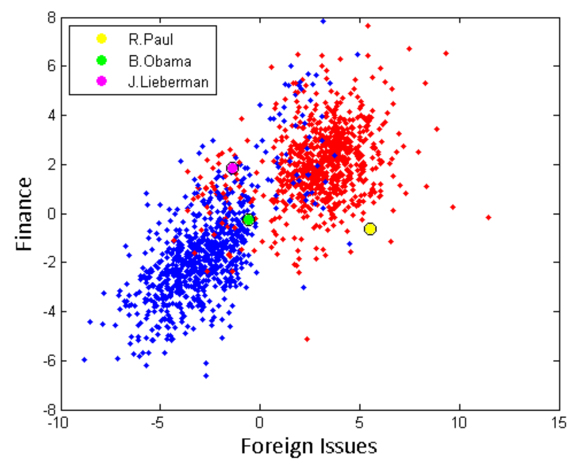


Figure 16: Scatter plot between "foreign issue" and "finance".

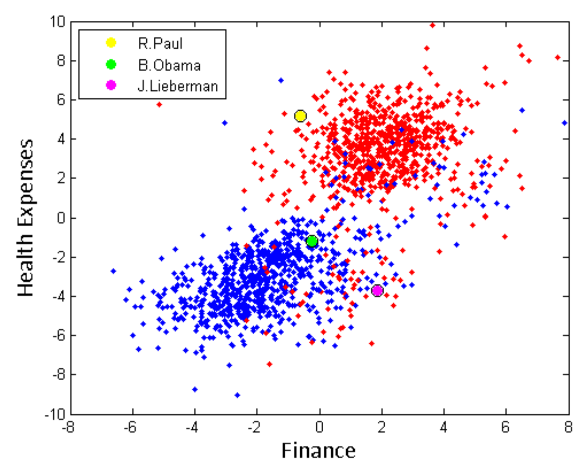


Figure 17: Scatter plot between "finance" and "health expenses".

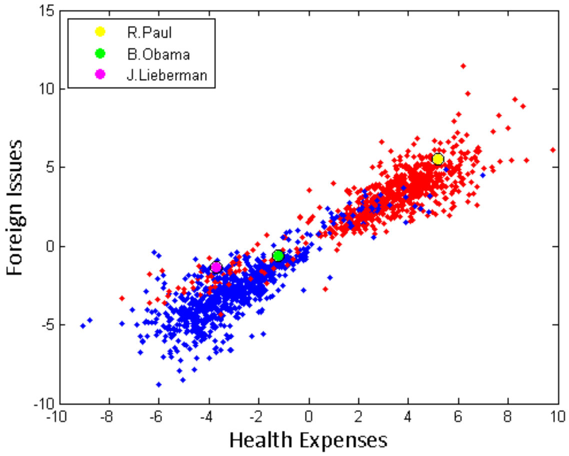


Figure 18: Scatter plot between "health expenses" and "foreign issue".

4.4 Application: Predicting Votes for Unseen Bills

An interesting application of our model is to make voting predictions for an unseen bill d with just text information. Note that, for 1-IPM and H-IPM, there is no way for us to predict votes related to an unseen bill, as it is impossible to get the latent feature for an unseen bill.

We perform our study in the following three steps:

- First, we estimate the topic distribution for the bill $\boldsymbol{\theta}_d$ according to current topic models, i.e., the $\boldsymbol{\beta}$, using simple language models.
- Second, we train a linear regression model to predict $\boldsymbol{a}_d$ and b_d, utilizing the text information (bag-of-word representation) as features. Note that, once we have learned the ideal points and popularity for the observed bills, we can then use these bills as training samples to learn the weights associated with the linear model. Using text to analyze data actually belongs to a broader study of opinion mining [16]. Here, we just use a simple approach to demonstrate our TF-IPM model.
- Finally for each legislator u, we can predict whether she favors this unseen bill by plugging in her ideal points in each topic: $p(v_{ud} = 1) = \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d)$.

We find that by using 90% of the bills as training dataset and hide all the voting information for the remaining 10% bills as testing dataset, we can still achieve a 80.8% accuracy in predicting votes from voters to them. More studies along this line will be performed in future work.

5. RELATED WORK

In this section, we introduce other related work to our study.

5.1 Ideal Point Models

Poole and Rosenthal ([17], [18], [19]) make seminal contributions and launched a massive literature of ideal point models. Alternative methods of ideal point estimation are developed by political scientists include Clinton et al. [4], Heckman and Snyder [8], and Londregan [12]. In [7], an issue-adjusted ideal point model is developed in order to evaluate a legislator's position for each explicit topic in her voting behavior. Apart from a global position, legislators have adjusted positions for different topics. They try to explain legislators' voting behavior by their affinity to topics that a bill covers. A network-based method is proposed to estimate ideal points for Twitter users in [2]. The key idea is that Twitter users prefer to follow politicians whose political positions are similar to theirs. People's political positions are then inferred by those whom they follow in social network.

5.2 Recommendation System

Ideal point estimation model is closely related to recommendation tasks in computer science domain. Given a user-item rating matrix, where the entry at i^{th} row and j^{th} column denotes the rating score from user i to item j, the recommendation task is to predict the rating score of items for users. Traditional recommendation algorithms include collaborative filtering (CF) [9] and latent factor models [11, 15].

In latent factor models, people try to explain ratings by characterizing users and items to several latent topics. This idea is natural in that, people usually rate an item based on several aspects, and items also have the corresponding properties. One of the most well known latent factor models is matrix factorization [11]. In matrix factorization models, each user i is represented by a vector $\mathbf{u}_i$ and item by a vector $\mathbf{v}_j$ in the K-dimensional latent space. The rating of user i to item j can be seen as the inner-product of the user feature vector and the item feature vector: $\hat{r}_{ij} = \mathbf{u}_i \mathbf{v}_j^T$. It naturally leads to the decomposition of the rating matrix R into the product of two matrices $R = UV^T$, where U is an $N \times K$ user-factor matrix, and V is an $M \times K$ item-factor matrix.

In this paper, we focus on the modeling of the voting prediction on a legislative voting network, where legislators do not have freedom to choose bills to vote. Also, we want to include the text information to enhance the prediction accuracy.

5.3 Topic Modeling, Link-Enhanced Topic Modeling, and Text-Enhanced Recommendation

Topic model is a type of statistic model for discovering the abstract topics that occur in a collection of documents. In topic models, each document is represented as a bag of words. The basic idea is, a latent topic is drawn conditionally to the document, and each word is then generated from that topic. The classical topic models include PLSA [10] and LDA [3].

In addition to the text information in documents, the links between documents can be exploited to further enhance the result of topic models. [20] proposes a unified information network-enhanced topic model that integrates the structural information and text information in a document network, and [14] proposes a solution to the problem of topic modeling with network structure regularization. Both of the two studies can significantly enhance the original topics model quality. In this paper, the topic model part is enhanced from the voting link part, which is different from the two papers where links exist between the documents.

Topic model can be used to enhance recommendation as well [7, 1, 22]. A collaborative topic regression model is proposed to recommend scientific articles [21]. The recommendation algorithm is based on both content of articles and users' ratings, and topic distributions are converted into feature vectors for users. Different from these methods, we use topic distribution as a explicit dimension for ideal points analysis, while existing methods directly use topic distribution as feature vector for users or items. In [13], a topic model is built on the reviews from users to products to help the rating prediction. However, the text information exists directly on the links from users to items, while in our case, text information is only linked to bills (items).

6. CONCLUSION

In this paper, we propose a novel topic-factorized ideal point estimation model (TF-IPM) that uses text information and voting behavior collectively in a legislative voting network. First, we model the ideal points of legislators and bills for each topic instead of assigning them to a global one. Second, the generation of topics are guided by the voting matrix in addition to the text information contained in bills. A unified model that combines voting behavior modeling and topic modeling is presented, and an iterative learning algorithm is proposed to learn the topics of bills as well as the topic-factorized ideal points of legislators and bills. By comparing with the state-of-the-art ideal point estimation models, our method has a much better explanation power in terms of held-out log-likelihood and other measures. Besides, case studies show that the topic-factorized ideal points coincide with human intuition. Finally, we illustrate how to use these topic-factorized ideal points to predict voting results for unseen bills.

7. REFERENCES

[1] D. Agarwal and B.-C. Chen. flda: matrix factorization through latent dirichlet allocation. In *Proc. of the third ACM Int. Conf. on Web Search and Data Mining (WSDM'10)*, pages 91–100, 2010.

[2] P. Barberá. Birds of the same feather tweet together. bayesian ideal point estimation using twitter data. *Unpublished manuscript, New York University*, 2013.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of Machine Learning Research (JMLR'03)*, 3:993–1022, 2003.

[4] J. Clinton, S. Jackman, and D. Rivers. The statistical analysis of roll call data. *American Political Science Review*, 98(02):355–370, 2004.

[5] S. Feldman and C. Johnston. Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, 2013.

[6] S. Gerrish and D. M. Blei. Predicting legislative roll calls from text. In *Proc. of the 28th Int. Conf. on Machine Learning (ICML'11)*, pages 489–496, 2011.

[7] S. Gerrish and D. M. Blei. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems (NIPS'12)*, pages 2762–2770, 2012.

[8] J. J. Heckman and J. M. S. Jr. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *The RAND Journal of Economics*, 28:pp. S142–S189, 1997.

[9] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proc. of the 22nd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 230–237, 1999.

[10] T. Hofmann. Probabilistic latent semantic analysis. In *Proc. of the Fifteenth Conf. on Uncertainty in Artificial Intelligence (UAI'99)*, pages 289–296, 1999.

[11] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[12] J. Londregan. Estimating legislators' preferred points. *Political Analysis*, 8(1):35–56, 1999.

[13] J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proc. of the 7th ACM Conf. on Recommender Systems (RecSys'13)*, pages 165–172, 2013.

[14] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In *Proc. of the 17th Int. Conf. on World Wide Web (WWW'08)*, pages 101–110, 2008.

[15] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems (NIPS'07)*, pages 1257–1264, 2007.

[16] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[17] K. T. Poole and H. Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):pp. 357–384, 1985.

[18] K. T. Poole and H. Rosenthal. Patterns of congressional voting. *American Journal of Political Science*, 35(1):pp. 228–278, 1991.

[19] K. T. Poole and H. Rosenthal. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, 1997.

[20] Y. Sun, J. Han, J. Gao, and Y. Yu. iTopicModel: Information network-integrated topic modeling. In *Proc. 2009 Int. Conf. Data Mining (ICDM'09)*, Miami, FL, 2009.

[21] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proc. of the 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'11)*, pages 448–456, 2011.

[22] E. Wang, D. Liu, J. Silva, L. Carin, and D. B. Dunson. Joint analysis of time-evolving binary matrices and associated documents. In *Advances in Neural Information Processing Systems (NIPS'10)*, pages 2370–2378, 2010.

[23] C. Zhai, A. Velivelli, and B. Yu. A cross-collection mixture model for comparative text mining. In *Proc. of the tenth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'04)*, pages 743–748, 2004.