
Abstract

We introduce the Hierarchical Ideal Point Topic Model, which provides a rich picture of policy issues, framing, and voting behavior using a joint model of votes, bill text, and the language that legislators use when debating the bills. We use this model to look at the relationship between Tea Party Republicans and “establishment” Republicans in the US House of Representatives during the 112th Congress.

1 Introduction

Ideal-point models are one of the most widely used tools in contemporary political science research (Poole and Rosenthal, 2007). These models estimate political preferences for legislators, known as their ideal points, from binary data such as legislative votes. Popular formulations analyze legislators’ votes and then place them on a one-dimensional scale, which is most often interpreted as an ideological spectrum from liberal to conservative (Martin and Quinn, 2002; Bafumi et al., 2005; Gerrish and Blei, 2011).

Moving beyond a single liberal-to-conservative dimension is attractive, however, since people may lean differently on different policy issues; for example, the conservative movement in the U.S. includes some fiscal conservatives who are relatively liberal on social issues, and vice versa. In multi-dimensional ideal point models, therefore, the ideal point of each legislator is no longer characterized by a single number, but by a multi-dimensional vector (Heckman and Jr., 1997; Jackman, 2001; Clinton et al., 2004). With that move comes a new challenge, though: the additional dimensions are often difficult to interpret. To mitigate this problem, recent research has introduced methods that estimate multi-dimensional ideal points using both voting data and the texts of the bills being voted on, e.g. using topic models and associating each dimension of the ideal point space with a topic (Gerrish and Blei, 2012; Lauderdale and Clark, 2014; Gu et al., 2014; Sim et al., 2015). The words most strongly associated with the topic can sometimes provide a readable description of its corresponding dimension.

In this paper, we develop this idea further by introducing HIPTM, the Hierarchical Ideal Point Topic Model, to estimate multi-dimensional ideal points for legislators in the US Congress. HIPTM differs from previous models in three key ways. First, HIPTM uses not only votes and associated bill text, but also the language of the legislators themselves. Intuitively, what a legislator chooses to talk about (agenda setting), and how they talk about it (framing), should be useful indicators of their underlying perspectives and preferences. Second, HIPTM improves the interpretability of ideal-point dimensions by incorporating data from the Congressional Bills Project (http://www.policyagendas.org/), in which bills are labeled with major topic headings from the widely adopted Policy Agendas Project Topic Codebook. And third, HIPTM discovers a hierarchy of topics, which allows us to analyze both agenda issues and issue-specific frames that legislators use on the congressional floor, following Nguyen et al. (2013) in modeling framing as second-level agenda setting (McCombs, 2005).

Using this new model, we focus on Republican legislators during the 112th US Congress, from January 3, 2011 until January 3, 2013. This is a particularly interesting session of Congress for political scientists, because of the emergence of the Tea Party, a decentralized political movement with populist, libertarian, and conservative elements, as a major force in American politics. Although united with “establishment” Republicans against

---

1 Ideal points are also used in other settings, such as judicial decision-making [CITE?].

2 http://www.policyagendas.org/
the Democrats in the 2010 midterm elections, leading to massive Democratic defeats, the Tea Party was — and still is — wrestling with establishment Republicans for control of the Republican party.

The Tea Party is a new and complex phenomenon for political scientists; as Carmines and D’Amico (2015) observe: “Conventional views of ideology as a single-dimensional, left-right spectrum experience great difficulty in understanding or explaining the Tea Party.” It is, therefore, a promising area to explore using new, multi-dimensional ideal-point models. HIPTM makes it possible to investigate a number of questions of interest to political scientists. For example, are there Republicans who identify themselves as members of the Tea Party, but whose votes and language betray a lack of enthusiasm for Tea Party issues? How well can we predict from someone’s language alone whether they are likely to associate themselves with the Tea Party? How do establishment and Tea Party Republicans differ in their framing of key issues under debate?

2 Related Work

Ideal point models (Carroll et al., 2009) are an application of item response theory (IRT), which describes probabilistic relationships between observed responses on a set of items by a set of responders who are characterized by some continuous latent traits (Fox, 2010). A popular formulation posits an ideal point \( u_a \) for each lawmaker \( a \), a polarity \( x_b \), and popularity \( y_b \) for each bill \( b \), all being values in \([-\infty, +\infty]\) (Martin and Quinn, 2002; Bafumi et al., 2005; Gerrish and Blei, 2011). Lawmaker \( a \) votes “Yes” on bill \( b \) with probability

\[
p(v_{a,b} = Y | u_a, x_b, y_b) = \Phi(u_a x_b + y_b)
\]

where \( \Phi(\alpha) = \exp(\alpha) / (1 + \exp(\alpha)) \) is the logistic (or inverse-logit) function.\(^3\) Intuitively, most lawmakers will vote “Yes” on bills with high popularity \( y_b \) and “No” on bills with low \( y_b \). When a bill’s popularity is near zero, the outcome of \( v_{a,b} \) depends on the interaction between the lawmaker’s ideal point \( u_a \) and the bill’s polarity \( x_b \).

In one-dimensional ideal point models, if two legislators have similar ideal points, they will vote similarly on every bill. To allow for different preferences on different policy issues, multi-dimensional ideal point models replace scalars \( u_a \) and \( x_b \) with \( K \)-dimensional vectors \( u_a \) and \( x_b \) (Heckman and Jr., 1997; Jackman, 2001; Clinton et al., 2004). The probability of Yes then becomes

\[
p(v_{a,b} = Y | u_a, x_b, y_b) = \Phi\left( \sum_{k=1}^{K} u_{a,k} x_{b,k} + y_b \right)
\]

Unfortunately, as Lauderdale and Clark (2014) observe, the binary data most commonly used for these models tend to be “insufficiently informative to support analyses beyond one or two dimensions”, and the additional dimensions are difficult to interpret. As a result, recent research has proposed adding topic models such as LDA into the mix. For example, Gerrish and Blei (2012) introduce the issue-adjusted ideal point model, which posits that each legislator \( a \) is characterized by a base ideal point \( u_a \) and an issue-adjusted vector \( z_a \), defining

\[
p(v_{a,b} = Y | u_a, z_a, x_b, y_b, w_b) = \Phi\left( \sum_{k=1}^{K} z_{a,k} \theta_{b,k} + u_a x_b + y_b \right)
\]

where \( \theta_{b,k} \) denotes the topic proportion of bill \( b \) estimated from its text \( w_b \). Other variations include the topic-factorized ideal point model (Gu et al., 2014) and Lauderdale and Clark’s (2014) vote-specific ideal points. Sim et al. (2015) use a similar framework to study judicial decisions of the US Supreme Court, incorporating amici curiae (“friends of the court” briefs) from non-litigants who weigh in on the decision. Other explorations include computing ideal points using campaign contributions (Bonica, 2013; Bonica, 2014) and ideal point estimation using Twitter data (Barberá, 2015).

3 Hierarchical Ideal Point Topic Model

In this section, we describe how we create a model that can discover frame-specific ideal points. Unlike previous work that discovers topic-specific ideal points (Gerrish and Blei, 2011), we can discover how politicians framing of issues are reflected (or not) in their votes.

Our model’s inputs are votes \( \{v_{a,b}\} \), each the response of \( a \in [1, A] \) to item \( b \in [1, B] \). Two types of text supplement the votes: (1) \( D \) speeches \( \{w_d\} \) from voter \( a_d \) and \( B \) descriptions \( \{w_i^b\} \) of bill \( b \). Even though congressional speeches are usually about a certain bill or a collection of related bills, we make no assumptions about the mapping
between \( w_d \) and \( w'_d \). This allows \( w'_d \) to be text by legislator \( a_d \) from blogs, social media, press releases etc. Figure 1 shows the plate notation diagram of the HIPTM, which has the following generative process:

1. For each issue \( k \in [1, K] \)
   (a) Draw a global distribution over frames \( \psi_k \sim \text{GEM}(\lambda_0) \)
   (b) Draw a topic \( \phi_k \sim \text{Dir}(\beta, \phi_0^k) \)
   (c) For each frame \( j \in [1, \infty] \)
      i. Draw a topic \( \phi_{k,j} \sim \text{Dir}(\beta, \phi_0^k) \)
      ii. Draw regression weight \( \eta_{k,j} \sim \mathcal{N}(0, \gamma) \)
2. For each document \( d \in [1, D] \)
   (a) Draw a topic proportion \( \theta_{d,k} \sim \text{Dir}(\alpha) \)
   (b) For each issue \( k \in [1, K] \), draw frame distribution \( \psi_{d,k} \sim \text{DP}(\lambda, \psi_0) \)
   (c) For each token \( t \in [1, N_d] \)
      i. Draw an issue \( z_{d,t} \sim \text{Mult}(\theta_d) \)
      ii. Draw a frame given the issue \( z_{d,t} \sim \text{Mult}(\psi_{d,z_{d,t}}) \)
      iii. Draw word \( w_{d,t} \sim \text{Mult}(\phi_{z_{d,t},m}) \)
3. For each voter \( a \in [1, A] \) on each issue \( k \in [1, K] \)
   (a) Draw issue’s ideal point \( u_{a,k} \sim \mathcal{N}(\sum_{j=1}^J \psi_{a,k,j} \eta_{k,j}, \rho) \)
4. For each bill \( b \in [1, B] \)
   (a) Draw polarity \( x_b \sim \mathcal{N}(0, \sigma) \) and popularity \( y_b \sim \mathcal{N}(0, \sigma) \)
   (b) Draw topic proportion \( \theta_b \sim \text{Dir}(\alpha) \)
   (c) For each token \( t \in [1, M_b] \)
      i. Draw an issue \( z'_{t,b} \sim \text{Mult}(\theta_b) \)
      ii. Draw a word type \( w'_{d,t} \sim \text{Mult}(\phi_{b,m}) \)
5. For each vote \( v_{a,b} \) of voter \( a \) on bill \( b \)
   (a) \( p(v_{a,b} = 1 | u_{a,b}, x_b, y_b, \theta_b, \phi_b) = \Phi \left( x_b \sum_{k=1}^K \phi_{k} u_{a,k} + y_b \right) \)

### 3.1 Defining the Topic Hierarchy

With the goal of analyzing agendas and frames in mind, our topic hierarchy has two levels: (1) issue nodes and (2) frame nodes. More specifically, there are \( K \) issue nodes, each with a topic \( \phi_k \) drawn from a Dirichlet distribution with concentration parameter \( \alpha \) and a prior mean vector \( \phi_0^k \), i.e., \( \phi_k \sim \text{Dir}(\beta, \phi_0^k) \).

To improve topic interpretability, we tie issue nodes to known political issues. This connection is through empirical word distribution estimated from the Congressional Bills Project \{\phi_k^*\}. Table 1 shows the prior distribution \( \phi_k^* \) for selected issues \( k \).

<table>
<thead>
<tr>
<th>Issue Category</th>
<th>Example Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>food, agriculture, loan</td>
</tr>
<tr>
<td>Banking, Finance, and Domestic Commerce</td>
<td>insurance, bank, patent</td>
</tr>
<tr>
<td>Defense</td>
<td>unit, army, force</td>
</tr>
<tr>
<td>Energy</td>
<td>oil, electricity, fuel</td>
</tr>
<tr>
<td>Health</td>
<td>drug, medicine, coverage</td>
</tr>
<tr>
<td>Labor, Employment, and Immigration</td>
<td>employment, immigration</td>
</tr>
<tr>
<td>Social Welfare</td>
<td>social security, disable</td>
</tr>
<tr>
<td>Transportation</td>
<td>transport, highway</td>
</tr>
<tr>
<td>Other</td>
<td>.....</td>
</tr>
</tbody>
</table>

4. The Congressional Bills Project provides a large collection of labeled congressional bill text: we compute the empirical word distribution for each issue all bills labeled with \( k \).
an issue $z_{d,n}$ from a document-specific multinomial distribution $\theta_d$, (2) then choose a frame $d_{d,n}$ from the set of infinitely many possible frames of the given issue $z_{d,n}$ using the frame proportion $\psi_{d,k}$ drawn from a Dirichlet process, and (3) finally choose a word type from the chosen frame’s topic

\[ \phi_{d_{d,n},z_{d,n}}. \]

Like LDA, to define the prior distribution for the topic proportion $\theta_d$ of each speech $d$, we use a symmetric Dirichlet distribution $\theta_d \sim \text{Dir}(\alpha)$. For each issue $k \in [1, K]$, the document-specific distribution over frames $\psi_{d,k}$ is distributed according to a Dirichlet process $\text{DP}(\lambda, \psi_k)$ with $\lambda$ as the concentration parameter and the issue-specific global distribution $\psi_k$ as the base distribution. In other words, our model generates text in the speeches using a mixture of $K$ HDPs (Teh et al., 2006). If we abandon the labeled data from the Congressional Bills Project to obtain the prior means $\phi_k^*$ for the 19 topics, it is relatively straightforward to extend to a fully nonparametric model where $K$ is unbounded (Ahmed et al., 2013).

### 3.3 Generating Bill Text

The bill text provides information about the policy agenda issues that each bill addresses. We use LDA to model the bill text $\{w_b\}$. Each bill $b$ is a mixture $\theta_b$ over $K$ issues, which is again drawn from a symmetric Dirichlet prior, i.e., $\theta_b \sim \text{Dir}(\alpha)$. Each token $w_{b,m}^t$ in bill $b$ is generated by first choosing a topic $z_{b,m}^t \sim \text{Mult}(\theta_b)$, and then choosing a word type $w_{b,m}^t \sim \text{Mult}(\phi_{z_{b,m}^t})$, just like LDA’s generative process.

### 3.4 Generating Roll Call Votes

Following recent work on multi-dimensional ideal points (Lauderdale and Clark, 2014; Sim et al., 2015), we define the probability of legislator $a$ voting “Yes” on bill $b$ as $p(v_{a,b} = \text{Yes} | u_{a}, x_b, y_b, \hat{\theta}_b) = \Phi \left( x_b \sum_{k=1}^{K} \hat{\theta}_{b,k} u_{a,k} + y_b \right) \tag{1}$

where $\hat{\theta}_b$ is the empirical distribution of bill $b$ over the $K$ issues and is defined as $\hat{\theta}_{b,k} = \frac{M_{b,k}}{M_{b}}$. Here, $M_{b,k}$ is the number of times in which tokens in $b$ are assigned to issue $k$ and $M_{b}$ is the marginal count, i.e., the number of tokens in bill $b$.

The ideal point of legislator $a$ specifically on issue $k$ is $u_{a,k}$ and comes from a normal distribution

\[ \sim \mathcal{N}(\hat{\phi}_{a,k}^T \eta_k, \rho) \equiv \mathcal{N} \left( \sum_{j=1}^{J_k} \hat{\psi}_{a,k,j} \eta_{k,j}, \rho \right) \tag{2} \]

where $J_k$ is the number of frames for topic $k$, which is unbounded. The mean of the Normal distribution is a linear combination of the ideal points $\{\eta_{k,j}\}$ of all issue $k$’s frames, weighted by how much time legislator $a$ spends on each frame when talking about issue $k$, i.e., $\psi_{a,k,j} = \frac{N_{a,k,j}}{N_{a,k}}$. Here, $N_{a,k,j}$ is the number of tokens authored by $a$ that are assigned to frame $j$ of issue $k$, and $N_{a,k}$ is the marginal count. When $N_{a,k} = 0$, which means that legislator $a$ does not talk about issue $k$, we back off to an uninformed zero mean.

Equation 2 resembles how supervised topic models (SLDA) link topics with a response, in that the response, the issue-specific ideal points $u_{a,k}$, is latent. Like Gerrish and Blei (2011), we use the bill text to regress on the bill’s latent polarity $x_b$ and popularity $y_b$. We only use text from congressional speeches for regression, as these can capture how legislators frame specific topics. Incorporating the bill text into the regression as well is an interesting direction for future work.

### 4 Posterior Inference

Given observed data which consist of (1) a set of legislative votes $\{v_{a,b}\}$ by $A$ legislators on $B$ bills, (2) a collection of congressional speeches $\{w_b\}$, each of which is given by a legislator $a_d$ and (3) the bill text $\{w_b^t\}$, we estimate the posterior distributions over the latent variables in our model using a stochastic EM inference algorithm. We alternate between (1) sampling the issue assignments $\{z_{b,m}\}$ for tokens in the bill text, (2) sampling the issue assignments $\{z_{d,n}\}$ and frame assignments $\{t_{d,n}\}$ for tokens in the speeches, (3) sampling the topics at first-level issue nodes $\{\phi_k\}$, (4) sampling the global frame proportion $\{\psi_k\}$ for all issues, (5) optimizing frames’ regression parameters $\{\eta_{k,j}\}$ using L-BFGS (Liu and Nocedal, 1989), and (6) updating the legislators’ multi-dimensional ideal points $\{u_{a,k}\}$ and the bills’ polarity $\{x_b\}$ and popularity $\{y_b\}$ using gradient ascent.

### Sampling Issue Assignments for Bill Tokens

The probability of assigning a token $w_{b,m}^t$ in the
bill text to an issue \( k \) is

\[
p(z_{b,m} = k \mid \text{rest}) \propto \frac{M_{b,k}^{-b,m} + \alpha}{M_{b,k}^{-b,m} + K \alpha} \cdot \hat{\phi}_{k, w_{b,m}} \tag{3}
\]

where \( M_{b,k} \) denotes the number of tokens in bill text \( b \) that are assigned to issue \( k \). The current estimated probability of word type \( v \) given issue \( k \) is denoted by \( \hat{\phi}_{k,v} \), which we update during the inference as described in Section 4. Marginal counts are denoted by \( \cdot \) and the superscript \(-b,m\) denotes the exclusion of the assignment for token \( w_{b,m} \) from the corresponding count.

**Sampling Frame Assignments in Speeches**

To sample the assignments for tokens in the speeches, we first sample an issue using the following sampling equation

\[
p(z_{d,n} = k \mid \text{rest}) \propto \frac{N_{d,n}^{-d,n} + \alpha}{N_{d,n}^{-d,n} + K \alpha} \cdot \hat{\phi}_{k, w_{d,n}} \tag{4}
\]

where \( N_{d,k} \) similarly denotes the number of times that tokens in \( d \) are assigned to issue \( k \). Given the sampled issue \( k \), we sample the frame as \( p(t_{d,n} = j \mid z_{d,n} = k, a_d = a, \text{rest}) \propto \mathcal{N}(u_{a,k}; \mu_{a,k,j}; \rho) \cdot \left( \frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j} \right)
\]

\[
\mathcal{N}(u_{a,k}; \mu_{a,k,j}; \rho) \cdot \left( \frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j} \right)
\]

\[
\frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j}
\]

\[
\frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j}
\]

\[
\frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j}
\]

\[
\frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j}
\]

\[
\frac{N_{d,n}^{-d,n} + \lambda}{N_{d,n}^{-d,n} + \lambda} \cdot \hat{\psi}_{k, j}
\]

where \( \mu_{a,k,j} = (\sum_{j'=1}^{J_k} \eta_{k,j'}N_{d,k,j'} + \eta_{k,j})/N_{d,k} \) for an existing frame \( j \), and for a newly created frame \( j' \), we have \( \mu_{a,k,j'} = (\sum_{j'=1}^{J_k} \eta_{k,j'}N_{d,k,j'} + \eta_{k,j})/N_{d,k} \). Following the Gaussian prior \( \mathcal{N}(0, \gamma) \). Here, the estimated global probability of choosing a frame \( j \) of issue \( k \) is \( \hat{\psi}_{k, j} \). We describe how we update this probability during inference in Section 4.

**Sampling Issue Topics**

In the generative process of HIPTM, the topic \( \phi_k \) of issue \( k \) (1) generates tokens in the bill text and (2) provides the Dirichlet priors of the issue’s frames. Following (Ahmed et al., 2013), we sample

\[
\hat{\phi}_k \sim \text{Dir}(m_k + \tilde{n}_k + \beta \phi_k^*) \tag{6}
\]

where \( m_k \equiv (M_{k,1}, M_{k,2}, \ldots, M_{k,V}) \) is the token count vector from the bill text assigned to each issue. The vector \( \tilde{n}_k \equiv (\tilde{N}_{k,1}, \tilde{N}_{k,2}, \ldots, \tilde{N}_{k,V}) \) denotes the token counts propagated from words assigned to topics that are associated with frames of issue \( k \), which can be approximated using the minimal or maximal path assumptions (Cowans, 2006).

**Sampling Frame Proportions**

Following the direct assignment method described in (Teh et al., 2006), we sample the global frame proportion

\[
\psi_k \equiv (\hat{\psi}_{k,1}, \hat{\psi}_{k,2}, \ldots, \hat{\psi}_{k,J_k}) \sim \text{Dir}((\tilde{N}_{k,1}, \tilde{N}_{k,2}, \ldots, \tilde{N}_{k,J_k}, \lambda_0) \tag{7}
\]

where \( \tilde{N}_{k,j} = \sum_{d=1}^{D} \tilde{N}_{d,k,j} \) and \( \tilde{N}_{d,k,j} \) can be sampled effectively using the Antoniak distribution (Antoniak, 1974). More details can be found in (Teh et al., 2006, page 1574).

**Optimizing Frame Regression Parameters**

We update the regression parameters \( \eta_k \) of frames under issue \( k \) using L-BFGS (Liu and Nocedal, 1989) to optimize \( \mathcal{L}(\eta_k) \)

\[
-\frac{1}{2\rho} \sum_{a=1}^{A} (u_{a,k} - \eta_k^T \hat{\psi}_{a,k}) - \frac{1}{2\gamma} \sum_{j=1}^{J_k} \eta_{k,j}^2 \tag{8}
\]

**Updating Ideal Points, Polarity and Popularity**

We update the multi-dimensional ideal point \( u_a \) of each legislator \( a \) and the polarity \( x_b \) and popularity \( y_b \) of each bill \( b \) by optimizing the log likelihood using gradient ascent.

**5 Predicting Tea Party Membership**

To quantitatively evaluate the effectiveness of our proposed HIPTM model in capturing “Tea Partisanship”, we predict Tea Party Caucus membership of legislators given votes and text. This examines (1) how effective the baseline features extracted from the votes and text are in predicting the Caucus membership, and (2) how much prediction improves using features from HIPTM.

For baselines, we consider simple feature sets:

- **Normalized term frequency (TF):** each legislator is a vector of term frequency of all word types in the vocabulary, normalized to unit length.
- **TF-IDF:** each legislator is represented by a TF-IDF vector.
- **VOTE:** each legislator is represented by a binary vector containing their voting record on the set of key votes selected by Freedom Works.

In our dataset from the 112th US Congress, there are 240 Republican Representatives. Sixty self-identify as Tea Party Caucus members. Our data are divided using five-fold cross-validation using stratified sampling, which preserves the ratio of the
two classes in both the training and test sets. We use AUC-ROC, which measures the area under the Receiver-Operating-Characteristic (ROC) curve, as the evaluation metric. We use SVM$^{light}$ (Joachims, 1999) as our classifier. After preprocessing, our vocabulary contains 5,349 unique word types.

**Membership from Votes and Text** First, given the votes and text of all the legislators, we run HIPTM for 1,000 iterations with a burn-in period of 500 iterations. After burning in, we keep the sampled state of the model after every 50 iterations. The feature values are obtained by averaging over the 10 stored models. Each legislator $a$ is represented by a vector concatenating the following features:

- $K$ dimensional ideal point vector estimated from both votes and text $u_{a,k}$
- $K$ dimensional vector, estimating the ideal point using only text $\eta_k^T \psi_{a,k}$
- $B$ probabilities estimating $a$’s votes on $B$ bills $\Phi(\{x_b \sum_{k=1}^K \hat{\theta}_{b,k} u_{a,k} + y_b\})$

Figure 2 shows AUC-ROC results for our feature sets. All feature sets improve upon random prediction’s AUC-ROC 0.5. VOTE-based features clearly outperform text-based features like TF and TF-IDF. Combining VOTE with either TF or TF-IDF does not improve the prediction performance much (i.e., VOTE-TF and VOTE-TF-IDF). Features extracted from our model, HIPTM, also outperforms TF and TF-IDF significantly, but only slightly better than VOTE. However HIPTM and VOTE together significantly outperform VOTE alone.

**Membership Prediction from Text Only** In the previous section, we experiment with features from both the votes and the text from legislators to predict their Tea Party Caucus memberships. However, its applicability is limited since we need to have both the votes and text to be able to make predictions. In this section, we look at a more difficult, yet more practical problem, which predicts the Tea Party Caucus membership using only the text of new lawmakers.

We first run our inference algorithm on the training data, which includes both votes and text. After training, using multiple models, we sample the issue and frame assignments for each token of the text authored by test lawmakers. Since the votes are not available, in this section, HIPTM’s extracted features only consist of (1) the $K$ dimensional vector estimating legislators’ ideal point using text only $\eta_k^T \psi_{a,k}$, and (2) the $B$ probabilities estimating the votes $\Phi(\{x_b \sum_{k=1}^K \hat{\theta}_{b,k} u_{a,k} + y_b\})$.

Figure 3 compares our features with the two text-based baselines TF and TF-IDF. Since HIPTM can no longer access the votes in the test data, its performance drops significantly compared with VOTE. However, HIPTM still outperforms the two text-based baselines TF and TF-IDF. Thus, our model provides an effective set of features compared with other commonly used text-based baselines to capture the “Tea Partiness” of legislators.

6 Analyzing Tea Party Ideal Points

In this section, we examine legislator’s ideal points. We first expose Tea Party-specific ideal points by examining one-dimensional ideal points. However, this analysis is limited, and we move on to the

---

5We use the default settings of SVM$^{light}$, except that we set the cost-factor equal to the ratio between the number of negative examples (i.e., number of non-Tea Party Caucus members) and the number of positive examples (i.e., number of Tea Party Caucus members).

6The MCMC configuration is the same as in Section 5.
issue-specific ideal points that HIPTM enables.

### 6.1 Data Collection

Recall that our goal is to see the issue-specific ideal points of Tea Partiers in the US House of Representatives; but what makes a Tea Partier? To address that question, we use key votes identified by Freedom Works as the most important votes on issues of economic freedom. Led by former House Majority Leader Dick Armey (R-TX), Freedom Works is a conservative non-profit organization which promotes “Lower Taxes, Less Government, More Freedom” and has been widely associated with the Tea Party movement.7 Karpowitz et al. (2011) report that, among the endorsements of various Tea Party organizations, Freedom Works endorsements get out the votes for the Republican candidates in the 2010 midterms.

For the 112th Congress, Freedom Works selected 60 key votes, 40 in 2011 and 20 in 2012. We are interested in ideal points with respect to the Tea Party movement, i.e., on the anti-pro Tea Party dimension, we consider whether a legislator agrees with the position of Freedom Works on a bill the binary response used in scaling the ideal points. More specifically, we assign \( v_{a,b} \) to be 1 if legislator \( a \) agrees with Freedom Works on bill \( b \), and 0 otherwise. In addition to the votes, we obtained the bill text with labels and the congressional speeches from the Congressional Bills Project. In total, we have 240 Republicans, 60 who self-identify with the Tea Party Caucus, and 13,856 votes.

### 6.2 One-dimensional Ideal Points

First, as a baseline for comparison, we estimate the one-dimensional ideal points of each legislator in our dataset using Equation 2. Figure 4 shows the box plots of estimated Tea Party ideal points for both members and non-members of the Tea Party Caucus among Republican Representatives in the 112th US House. The median of members’ ideal points is significantly higher than that of non-members’ ideal points, though there is significant overlap.

![Figure 4: Box plots of the estimated one-dimensional Tea Party ideal points for members and non-members of the Tea Party Caucus among Republican Representatives in the 112th US House.](image)

From our estimate, Jeff Flake (R-AZ) has the second highest ideal point, but is not a member of the Tea Party Caucus. Looking more closely into his voting record, out of 60 key votes selected by Freedom Works he only disagrees with Freedom Works’s position on one where he voted “Nay” on the bill “H.R.1: Full-Year Continuing Appropriations Act, 2011”. This bill includes the largest single discretionary spending cut in history, cutting $106 billion from various programs and departments. Another example is Justin Amash (R-MI), who founded and is the Chairman the Liberty Caucus; its members are conservative and libertarian Republicans. Amash has agreed with Freedom Works on every single key votes selected by Freedom Works since 2011.

Conversely, there are members of the Tea Party Caucus who do not often agree with Freedom Works, and thus have relatively low ideal points. For example, Rodney Alexander (R-LA), who agrees with Freedom Works only 48% of the time in the 112th Congress, was a member of the Democrat party before changing his party affiliation in 2004. Another example is Ander Crenshaw (R-FL) with 50% agreement with Freedom Works’s positions on key votes in 2011 and 2012. Both Alexander and Crenshaw are categorized as “Green Tea” by (Gervais and Morris, 2014), which refers to Republican legislators who are strongly “associated with the Tea Party on their own initiative” but are not strongly supported by Tea Party organizations.

---


8 Estimated ideal point signs might be flipped, as \( u_a x_b = (-u_a)(-x_b) \), which makes no difference in Equation 2. To ensure that higher ideal points are “pro-Tea Party”, we first sort the legislators according to the fraction of votes for which they agree with Freedom Works and initialize the ideal points of the top and bottom five legislators with \(+3\sigma\) and \(-3\sigma\), where \(\sigma\) is the variance of \(u_a\)’s Gaussian prior.
6.3 Multi-dimensional Ideal Points

While it is interesting to compare holistic measures of Tea Partiness, it doesn’t reveal how legislators conform or deviate from what defines a mainstream Tea Partier. In this section, we use HIPTM to show how the issue-specific ideal points of the two groups of Republican Representatives differ.

Figure 5 shows the boxplots of the estimated ideal points for each policy agenda issue, sorted by the difference between the median of the two groups’ ideal points. On most issues, the ideal point distributions of the two Republican groups are nigh identical.

On several issues, the ideal point distributions of the two groups of legislators differ significantly. To understand why these issues polarize, we look at key votes. Recall that in our model, each bill $b$ has a distribution $\theta_b$ over $K$ issues, capturing what the bill is about. For each key vote $b$, we choose the issue with the highest probability $\theta_{bk}$ and use it to label. Although using a single issue for each key vote conflicts with our model’s admixture assumption in which each document is a mixture of topics, it provides a good estimation of what the key vote is primarily about and helps reduce the complexity of our analysis.

In the remainder of this section, we analyze the voting records of Republicans on the key votes that our model assigns to the “Government Operation”, “Macroeconomics”, and “Transportation” issues, which helps explain why our model estimates these issues as the most polarized.

Government operations The majority of the two groups, members vs. non-members of the Tea Party Caucus, vote differently on eight out of eleven key votes on this issue. Most of these key votes are to reduce the government spending on various federal programs including the Economic Development Administration (key vote 2012-207), the Energy Efficiency and Renewable Energy Program (key vote 2012-311) and the Fossil Fuel Research and Development programs (key vote 2012-317). More specifically, on the key vote to eliminate the Energy Efficiency and Renewable Energy Program (2012-311), nearly 80% (41 out of 53) of the Tea Party Caucus members vote “Yea” agreeing with the Freedom Works, while only about 43% of non-Tea Party Caucus members vote similarly. This difference in voting behaviors explains why this issue is estimated as the most polarized issue by our model, which aligns well with the agenda of the Tea Party movement fighting for less government and more federal spending cuts.

Macroeconomics Among Macroeconomics votes, there are two key votes where the two groups vote differently. These are key vote 2011–275 (To replace the Paul Ryan budget with the RSC’s budget) and key vote 2012–149 (Substitute amendment containing the Republican Study Committee budget for FY 2013). Both of these key votes, one in 2011 and the other in 2012, are to replace Paul Ryan’s budget plan with the Republican Study Committee’s (RSC) alternate proposal “Back to Basics” to cut government spending more aggressively to balance the federal budget in a decade. In 2011, the key vote 2011–275 split the Republican Representatives with 118 Yea’s and 119 Nay’s. However, while the majority (104 out of 119) of non-Tea Party Caucus members vote against this amendment, 45 out of 60 members of the Tea Party Caucus vote for it. In 2012, even more Tea Party Caucus members vote for the key
vote 2012–149, while there are still more than half of non-Tea Party Caucus members vote against it.

Although not as polarized as the two key votes above, the two remaining key votes still see disagreements among Republicans. The first key vote is to the *Budget Control Act of 2011* (key vote 2011–690) which allows President Obama to raise the debt ceiling to over $16 trillion, while the second key vote is about the *Taxpayer Relief Act of 2012* (key vote 2012–659) to avert the “fiscal cliff”.

**Transportation** “Transportation” is the third most polarized issue estimated by our model, with two key votes focusing on the federal spending on transportation. The first key vote (2012–378) caps highway spending at the amount taken in by the gas tax. More than half of Tea Party Caucus members (32 out of 55) vote for this motion, while non-members vote against it. Conversely, the second key vote (2012–451) authorizes federal highway spending at a level that far exceeds its revenue from the gas tax, which is opposed by Freedom Works.

7 Agendas and Frames: Analyzing Topic Hierarchy

TBD

8 Conclusion

In this paper, we introduced HIPTM, which integrates hierarchical topic modeling with multi-dimensional ideal points in order to jointly model voting behavior, the text content of bills, and the language used by legislators. The hierarchical structure of HIPTM follows Nguyen et al. (2013) in providing a well defined computational realization of agenda setting and framing, motivated by political science theory (McCombs, 2005). Moreover, its multi-dimensionality permits modeling of legislators as having different ideological leanings on different issues, and its leveraging of labeled data from the Congressional Bills Project enhances interpretability in order to facilitate analysis by political domain experts.

Looking at Republican votes and debates in the 112th US House of Representatives, analysis with single-dimensional ideal point modeling shows that Tea Party and non-Tea Party Republicans overlap significantly in ideological preference, which is unsurprising and not particularly informative. Moving to HIPTM, though, we discover a richer picture of the relationship between Tea Party and “establishment” Republicans, one that takes advantage of the hierarchy’s first level to illuminate the focus of Congressional floor debates, and the hierarchy’s second level to illuminate differences in framing by the two sometimes agreeing, sometimes conflicting groups of legislators.

Evaluating the model’s predictive power, we find that HIPTM is more effective than previous methods when applied to the task of predicting membership in the Tea Party Caucus. This task is introduced more as a well defined NLP evaluation than for its intrinsic interest; predicting caucus membership, and even vote prediction, *per se*, are not of particular interest to political scientists. However, it illustrates perhaps the most interesting aspect of the model: suggests the possibility of more effective ideal point modeling for new members of Congress who have not yet established a voting record. More intriguingly, it suggests the possibility of assessing the “Teapartyness” of candidates or even media outlets based on the language they use.

References


Royce Carroll, Jeffrey B Lewis, James Lo, Keith T Poole, and Howard Rosenthal. 2009. Measuring


Yanchuan Sim, Bryan Routledge, and Noah A Smith. 2015. The utility of text: The case of Amicus briefs and the Supreme Court. In *Association for the Advancement of Artificial Intelligence*.
