Guided Probabilistic Topic Models for Agenda-setting and Framing

Viet-An Nguyen

Ph.D. Dissertation Defense
Department of Computer Science

February 13, 2015

UNIVERSITY OF MARYLAND
Political agenda is the “set of issues that are the subject of decision making and debate within a given political system at any one time” [Baumgartner, 2001]
Political agenda is the “set of issues that are the subject of decision making and debate within a given political system at any one time” [Baumgartner, 2001]
Political agenda is the “set of issues that are the subject of decision making and debate within a given political system at any one time” [Baumgartner, 2001]
Political agenda is the “set of issues that are the subject of decision making and debate within a given political system at any one time” 
[Baumgartner, 2001]

WHAT do people talk about?
“To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described”

[Entman, 1993]
“To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described”

[Entman, 1993]
“To frame is to **select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described**”

[Entman, 1993]
“To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described”

[Entman, 1993]

How do people talk about certain issues?
## Content Analysis Approaches

<table>
<thead>
<tr>
<th>Approach</th>
<th>Costs [Quinn et al., 2010]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-analysis</td>
</tr>
<tr>
<td><strong>Manual content analysis approach</strong></td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td></td>
</tr>
<tr>
<td>Human coding</td>
<td></td>
</tr>
<tr>
<td><strong>Automated content analysis approach</strong></td>
<td></td>
</tr>
<tr>
<td>Supervised learning</td>
<td></td>
</tr>
<tr>
<td>Topic modeling</td>
<td></td>
</tr>
</tbody>
</table>
## Content Analysis Approaches

<table>
<thead>
<tr>
<th>Approach</th>
<th>Costs [Quinn et al., 2010]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-analysis</td>
</tr>
<tr>
<td>Manual content analysis approach</td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td></td>
</tr>
<tr>
<td>Human coding</td>
<td></td>
</tr>
<tr>
<td>Automated content analysis approach</td>
<td></td>
</tr>
<tr>
<td>Supervised learning</td>
<td></td>
</tr>
<tr>
<td>Topic modeling</td>
<td></td>
</tr>
</tbody>
</table>

### Costs

- **Pre-analysis cost:** incurred before the actual analysis happens
  - e.g., design codebook, train coders, and annotate data
- **Analysis cost:** incurred during the content analysis process
- **Post-analysis cost:** incurred after the analysis process
  - e.g., interpret analyzed results
## Content Analysis Approaches

<table>
<thead>
<tr>
<th>Approach</th>
<th>Costs [Quinn et al., 2010]</th>
<th>Pre-analysis</th>
<th>Analysis</th>
<th>Post-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manual content analysis approach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td>low</td>
<td>high</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Human coding</td>
<td>high</td>
<td>high</td>
<td>low</td>
<td></td>
</tr>
<tr>
<td><strong>Automated content analysis approach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervised learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Costs

- **Pre-analysis cost**: incurred before the actual analysis happens
  - e.g., design codebook, train coders, and annotate data
- **Analysis cost**: incurred during the content analysis process
- **Post-analysis cost**: incurred after the analysis process
  - e.g., interpret analyzed results

---

**Overview**
## Content Analysis Approaches

<table>
<thead>
<tr>
<th>Approach</th>
<th>Costs [Quinn et al., 2010]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-analysis</td>
</tr>
<tr>
<td><strong>Manual content analysis approach</strong></td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td>low</td>
</tr>
<tr>
<td>Human coding</td>
<td>high</td>
</tr>
<tr>
<td><strong>Automated content analysis approach</strong></td>
<td></td>
</tr>
<tr>
<td>Supervised learning</td>
<td>high</td>
</tr>
<tr>
<td><strong>Topic modeling</strong></td>
<td>low</td>
</tr>
</tbody>
</table>

**Figure**: Typical output of unsupervised topic models
## Content Analysis Approaches

<table>
<thead>
<tr>
<th>Approach</th>
<th>Costs [Quinn et al., 2010]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-analysis</td>
</tr>
<tr>
<td>Manual content analysis approach</td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td>low</td>
</tr>
<tr>
<td>Human coding</td>
<td>high</td>
</tr>
<tr>
<td>Automated content analysis approach</td>
<td></td>
</tr>
<tr>
<td>Supervised learning</td>
<td>high</td>
</tr>
<tr>
<td>Topic modeling</td>
<td>low</td>
</tr>
</tbody>
</table>

In this thesis

Following the topic modeling approach, we develop a series of new models, which are **guided by additional information associated with the text** and designed to discover and analyze agenda-setting and framing at a lower cost.
Main Contributions

**Goal:** Study agenda-setting and framing at a lower cost

1. Technical Contributions
2. Applications
3. Overview
Main Contributions

**Goal:** Study *agenda-setting* and framing at a *lower cost*

**Technical Contributions**

- Extend prior work on topic segmentation in conversation by incorporating *speaker identity* and using *Bayesian nonparametrics*
- *Speaker Identity for Topic Segmentation* (SITS)

**Applications**

- Study *agendas and agenda control* in political debates and other conversations
- Develop an *interactive visualization* to analyze results effectively
- Improve performance in topic segmentation and influencer detection
Main Contributions

**Goal:** Study *agenda-setting* and framing at a *lower cost*

**Technical Contributions**
- Capture **dependency among labels** in multi-labeled data using a **tree-structured topic hierarchy**
- *Label-to-Hierarchy* (L2H)

**Applications**
- Analyze **policy agenda issues** in legislative text and how they relate to each other
- Learn an **interpretable label hierarchy** to reduce post-analysis cost
- Improve performance in predicting held-out words and multiple labels of unseen documents
Main Contributions

**Goal:** Study *agenda-setting* and *framing* at a lower cost

### Technical Contributions

- Extend existing supervised topic model using a **hierarchy of topics**
- Combine topic regression with *lexical regression* to improve prediction
- *Supervised Hierarchical Latent Dirichlet Allocation* (SHLDA)

### Applications

- Provide a formal computational model corresponding to the theory of *framing as second-level agenda setting*
- Improve performance in ideology prediction and sentiment analysis
Main Contributions

**Goal:** Study agenda-setting and framing at a lower cost

---

**Technical Contributions**

- Extend existing **multi-dimensional ideal point** models using a **hierarchy of topics**
- **Hierarchical Ideal Point Topic Model** (HIPTM)

---

**Applications**

- Provide a formal computational model corresponding to the theory of **framing as second-level agenda setting**
- Analyze ideological positions of legislators in **multiple interpretable dimensions**
- Map frames onto issue-specific ideological dimensions which allows prediction about ideology using text only
Outline

1

2

3

4


Agenda Control Behaviors in 2008 Presidential Debates

- In presidential debates, moderators have much higher scores than candidates do.
- In the VP debate, IFILL’s score is only slightly higher than those of PALIN and BIDEN.

The Ifill Factor

By Scott Horton

.... Ifill's questioning and moderating was, as The Atlantic's James Fallows remarked, "terrible." She asked open-ended, utterly predictable questions which presented very little challenge to the candidates. But even more important to the McCain campaign's strategy, Palin was able to simply ignore the questions and recite her talking points.
Agenda Control Behaviors in 2008 Presidential Debates

- In presidential debates, moderators have much higher scores than candidates do.
- In the VP debate, IFILL’s score is only slightly higher than those of PALIN and BIDEN.

The Ifill Factor

By Scott Horton

.... Ifill's questioning and moderating was, as The Atlantic's James Fallows remarked, "terrible." She asked open-ended, utterly predictable questions which presented very little challenge to the candidates. But even more important to the McCain campaign's strategy, Palin was able to simply ignore the questions and recite her talking points.
Political Agenda in Legislative Texts

Policy Agenda Research in Political Science

What are the subjects of political attention?

Focus of much research in political science
- policy agenda change (Baumgartner and Jones 1993; Kingdon 1995, Quinn et al. 2010)
- issue evolution (Carmines and Stimson 1989; Wolbrecht 2000)
Approaches to Study Policy Agendas

Human coding

- Define codebook, train coders, annotate documents
  - Policy Agendas Project: define 19 major topics, 225 subtopics
  - Congressional Bills Project: one major topic for each bill

Unsupervised Topic modeling

- Unsupervised model to learn a set of topics, each of which is an agenda issue
## Approaches to Study Policy Agendas

### Human coding
- Define codebook, train coders, annotate documents
  - Policy Agendas Project: define 19 major topics, 225 subtopics
  - Congressional Bills Project: one major topic for each bill

Many bills are about more than one topic
- Difficult to extend over time and to other domains, e.g.,
  - “Immigration” was added to the Policy Agendas Codebook in 2014
  - “Arts and Entertainment”, “Churches and Religion” etc are added to analyze NY Times

### Unsupervised Topic modeling
- Unsupervised model to learn a set of topics, each of which is an agenda issue
## Approaches to Study Policy Agendas

### Human coding
- Define codebook, train coders, annotate documents
  - Policy Agendas Project: define 19 major topics, 225 subtopics
  - Congressional Bills Project: one major topic for each bill
- Many bills are about more than one topic
- Difficult to extend over time and to other domains, e.g.,
  - “Immigration” was added to the Policy Agendas Codebook in 2014
  - “Arts and Entertainment”, “Churches and Religion” etc are added to analyze NY Times

### Unsupervised Topic modeling
- Unsupervised model to learn a set of topics, each of which is an agenda issue
- Difficult to interpret outputs
### Multi-labeled Data

- Each document is tagged with **multiple labels** from a flexible, extendable vocabulary of labels.

<table>
<thead>
<tr>
<th>Documents</th>
<th>Multiple Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>International affairs</td>
</tr>
<tr>
<td></td>
<td>Military operations and strategy</td>
</tr>
<tr>
<td></td>
<td>Foreign aid and international relief</td>
</tr>
</tbody>
</table>
Analyzing Policy Agendas using Multi-labeled Data

Multi-labeled Data

- Each document is tagged with **multiple labels** from a flexible, extendable vocabulary of labels

**Pros:**
- Allow multiple labels per bill
- Avoid having to predefine a complete codebook
- Learn interpretable agenda issues

**Cons:**
- Number of labels is large → capture dependency among labels
- Labeling might not be exhaustive → handle missing labels

<table>
<thead>
<tr>
<th>Documents</th>
<th>Multiple Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>International affairs</td>
</tr>
<tr>
<td>Military operations and strategy</td>
<td>International affairs</td>
</tr>
<tr>
<td>Foreign aid and international relief</td>
<td>Middle East</td>
</tr>
</tbody>
</table>
Approach Overview

- Tree-structured hierarchy: captures label dependency and handles missing labels
- One topic per label: improves interpretability
**Approach Overview**

- **Tree-structured hierarchy**: captures label dependency and handles missing labels
- **One topic per label**: improves interpretability

---

- **International affairs**
  - **Military operations and strategy**
  - **Foreign aid and international relief**
  - **Int'l organizations & cooperation**
  - **International law and treaties**

---

- **Department of Defense**
- **Terrorism**
- **Asia**
- **Latin America**
- **Europe**
- **Middle East**
Approach Overview

- Tree-structured hierarchy: captures label dependency and handles missing labels
- **One topic per label**: improves interpretability

<table>
<thead>
<tr>
<th>Label (predefined)</th>
<th>Concept (label-topic pair)</th>
</tr>
</thead>
<tbody>
<tr>
<td>International affairs</td>
<td></td>
</tr>
<tr>
<td>libya, unit_nation, intern_religi, bahrain, religi_freedom, religi_minor, freedom_act, africa, violenc, secur_council, benghazi, privileg_resolut, hostil,</td>
<td></td>
</tr>
</tbody>
</table>

Label (predefined)  
Topic (learned)
L2H: Label-to-Hierarchy
L2H Generative Process
1. Generating topic tree

Construct a complete weighted directed graph $G$ where each label is a node and the edge weight is:

$$t_{\text{Health care}} \rightarrow \text{Medicare} = \frac{\text{No. docs tagged with Health care & Medicare}}{\text{No. docs tagged with Medicare}}$$

Add a Background node to the graph.

Generate a uniform spanning tree

$$p(T|G) = \prod_{\text{all edges (i} \rightarrow \text{j)}} t_{i \rightarrow j}$$

Associate each node with a topic:

$$\phi_{\text{Background}} \sim \text{Dir}(\beta u)$$
$$\phi_{\text{Medicare}} \sim \text{Dir}(\beta \phi_{\text{Health care}})$$

Political Agenda in Legislative Texts
Construct a complete weighted directed graph $G$ where each label is a node and the edge weight is:

$$t_{\text{Health care} \rightarrow \text{Medicare}} = \frac{\text{No. docs tagged with both Health care & Medicare}}{\text{No. docs tagged with Medicare}}$$
Construct a complete weighted directed graph $G$ where each label is a node and the edge weight is:

$$t_{\text{Health care} \rightarrow \text{Medicare}} = \frac{\text{No. docs tagged with both Health care & Medicare}}{\text{No. docs tagged with Medicare}}$$

- Add a *Background* node to the graph
L2H Generative Process

1. Generating topic tree

- Construct a complete weighted directed graph $\mathcal{G}$ where each label is a node and the edge weight is:

$$t_{\text{Health care} \rightarrow \text{Medicare}} = \frac{\text{No. docs tagged with both Health care & Medicare}}{\text{No. docs tagged with Medicare}}$$

- Add a Background node to the graph
- Generate a uniform spanning tree $p(\mathcal{T} | \mathcal{G}) = \prod_{\text{all edges } (i \rightarrow j)} t_{i \rightarrow j}$
L2H Generative Process
1. Generating topic tree

- Construct a complete weighted directed graph $G$ where each label is a node and the edge weight is:

  \[
  t_{\text{Health care} \rightarrow \text{Medicare}} = \frac{\text{No. docs tagged with both Health care & Medicare}}{\text{No. docs tagged with Medicare}}
  \]

- Add a Background node to the graph
- Generate a uniform spanning tree $p(T | G) = \prod_{\text{all edges } (i \rightarrow j)} t_{i \rightarrow j}$
- Associate each node with a topic:

  \[
  \begin{cases} 
  \phi_{\text{Background}} \sim \text{Dir}(\beta u) \\
  \phi_{\text{Medicare}} \sim \text{Dir}(\beta \phi_{\text{Health care}})
  \end{cases}
  \]
Given a document $d$ with labels $I_d$

- Define candidate set $L^1_d$ and the complementary set $L^0_d$
- For each token $n \in [1, N_d]$
  - Choose either $L^1_d$ or $L^0_d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $l_d$

- Define candidate set $\mathcal{L}_d^1$ and the complementary set $\mathcal{L}_d^0$
- For each token $n \in [1, N_d]$
  - Choose either $\mathcal{L}_d^1$ or $\mathcal{L}_d^0$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
L2H Generative Process
2. Generating documents

Given a document $d$ with labels $I_d$

- Define candidate set $\mathcal{L}_d^1$ and the complementary set $\mathcal{L}_d^0$
- For each token $n \in [1, N_d]$
  - Choose either $\mathcal{L}_d^1$ or $\mathcal{L}_d^0$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $I_d$
- Define candidate set $L^1_d$ and the complementary set $L^0_d$
- For each token $n \in [1, N_d]$
  - Choose either $L^1_d$ or $L^0_d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $I_d$

- Define candidate set $L^1_d$ and the complementary set $L^0_d$
- For each token $n \in [1, N_d]$
  - Choose either $L^1_d$ or $L^0_d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $I_d$

- Define candidate set $L^1_d$ and the complementary set $L^0_d$
- For each token $n \in [1, N_d]$
  - Choose either $L^1_d$ or $L^0_d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $I_d$

- Define candidate set $L^1_d$ and the complementary set $L^0_d$
- For each token $n \in [1, N_d]$
  - Choose either $L^1_d$ or $L^0_d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Given a document $d$ with labels $I_d$

- Define candidate set $L_1^d$ and the complementary set $L_0^d$
- For each token $n \in [1, N_d]$
  - Choose either $L_1^d$ or $L_0^d$
  - Choose a node in the chosen label set
  - Draw word from the node’s topic
Multi-label Classification

**Task:** Predicting multiple labels for each test document

**Evaluation Metric:** Macro-F1

**Bill H.R.62:** A bill to establish a series of six regional Presidential primaries at which the public may express its preference for the nomination of an individual for election to the Office of President of the United States.

**Data:** Text and labels from bills in 4 U.S. Congresses (109th-112th)
**Multi-label Classification**

**Task:** Predicting multiple labels for each test document

**Evaluation Metric:** Macro-F1

**Bill H.R.62:** A bill to establish a series of six regional Presidential primaries at which the public may express its preference for the nomination of an individual for election to the Office of President of the United States.

**Data:** Text and labels from bills in 4 U.S. Congresses (109th-112th)

<table>
<thead>
<tr>
<th>Congress</th>
<th>Number of bills</th>
<th>Number of labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>13067</td>
<td>418</td>
</tr>
<tr>
<td>110</td>
<td>14034</td>
<td>375</td>
</tr>
<tr>
<td>111</td>
<td>13673</td>
<td>243</td>
</tr>
<tr>
<td>112</td>
<td>12274</td>
<td>205</td>
</tr>
</tbody>
</table>
**Method:** Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features.
**Method:** Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features
Method: Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features.

![Graph showing Macro F1 scores for different feature sets: Term Frequency (TF), TF-IDF, L-LDA & TF-IDF, L2H & TF-IDF.](image-url)
Method: Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features

TF-IDF outperforms TF significantly
**Method:** Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features.
Method: Using M3L—an efficient max-margin multi-label classifier (Hariharan et al., MLJ’12) to study different sets of features

<table>
<thead>
<tr>
<th>Term Frequency (TF)</th>
<th>TF-IDF</th>
<th>L-LDA &amp; TF-IDF</th>
<th>L2H &amp; TF-IDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>110</td>
<td>111</td>
<td>112</td>
</tr>
<tr>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Macro F1

TF-IDF outperforms TF significantly
Adding L-LDA does not improve much over TF-IDF
Adding L2H improves TF-IDF significantly
TF-IDF outperforms TF significantly
Introduced L2H, a new hierarchical topic model for multi-labeled data, which

- captures label dependencies using a tree-based hierarchy
- provides an interpretable way to explore relationships between policy agenda issues
- improves multi-label classification performance
## Agenda-setting

<table>
<thead>
<tr>
<th>What gets talked about?</th>
</tr>
</thead>
<tbody>
<tr>
<td>~ Topics</td>
</tr>
</tbody>
</table>
Agenda-setting & Framing: Computational Approach

Agenda-setting

- **What** gets talked about?
- ~ Topics

Framing

- **How** things get talked about?
- ~ ???

[McCombs, 2004]
Agenda-setting & Framing: Computational Approach

Agenda-setting
- **What** gets talked about?
- ∼ Topics

Framing
- **How** things get talked about?
- ∼ ???

An approach: Framing = **Second-level agenda setting**
[McCombs, 2004]
Legalizing Marijuana

Economics

Health

Legal process

Framing as Second-level Agenda-setting
Input Data
A collection of documents $w_1, w_2, \cdots, w_D$

Document = debate turn

Each document $d$ has an associated response variable $y_d$

Response variable = ideological position (ideal point) of speaker on the liberal–conservative spectrum
Framing: Liberal vs. Conservative

Input Data

- A collection of documents $w_1, w_2, \ldots, w_D$
  - Document = debate turn
- Each document $d$ has an associated response variable $y_d$
  - Response variable = ideological position (ideal point) of speaker on the liberal–conservative spectrum
Topics are arranged in a tree-structured hierarchy
- High-level nodes: more general, map to agenda issues
- Low-level nodes: more specific, map to issue-specific frames

<table>
<thead>
<tr>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature</td>
</tr>
<tr>
<td>Externalities</td>
</tr>
<tr>
<td>Industry</td>
</tr>
<tr>
<td>Economy</td>
</tr>
<tr>
<td>Heath Care</td>
</tr>
<tr>
<td>Debates</td>
</tr>
</tbody>
</table>
Topics are arranged in a tree-structured hierarchy
- High-level nodes: more general, map to *agenda issues*
- Low-level nodes: more specific, map to *issue-specific frames*

Each node has a **regression parameter** specifying its position on the liberal–conservative spectrum

Environment
- Heath Care
  - R: 0.5
  - D: 0.9

Debates
- Nature
  - D: 1.4

Economy
- Externalities
  - R: 0.6

- Industry
  - R: 1.5
Modeling approach

- Topics are arranged in a tree-structured hierarchy
  - High-level nodes: more general, map to agenda issues
  - Low-level nodes: more specific, map to issue-specific frames
- Each node has a regression parameter specifying its position on the liberal–conservative spectrum
- What topics speakers talk about and what words they use will decide their ideological positions
For each node $k$ in the tree:
- Draw topic $\phi_k \sim \text{Dir}(\beta)$
For each node $k$ in the tree
- Draw topic $\phi_k \sim \text{Dir}(\beta)$
- Draw regression parameter $\eta_k \sim \mathcal{N}(\mu, \sigma)$
SHLDA: Generating words

Agenda-setting & Framing in Political Text
SHLDA: Generating words

- Each document is a bag of sentences

Diagram:
- Each document is a bag of sentences
- Sentences
- Debates
  - Heath Care: D: 0.9
  - Environment: D: 0.4
  - Economy: R: 0.5
- Nature: D: 1.4
- Externalities: R: 0.6
- Industry: R: 1.5

Diagram:
- Sentences
- Debates
  - Heath Care
  - Environment
  - Economy
- Nature
- Externalities
- Industry
SHLDA: Generating words

- Each document is a bag of sentences
- Each sentence is a bag of tokens
A Chinese restaurant process for each document to cluster similar sentences
Each sentence is assigned to a table
SHLDA: Generating words

- Each sentence is assigned to a table
- Each table is assigned to a path

![Diagram showing the assignment of sentences to tables and tables to paths.]

---

Agenda-setting & Framing in Political Text
Each sentence is assigned to a table
Each table is assigned to a path
Each token is assigned to a node on the chosen path
Each sentence is assigned to a table
Each table is assigned to a path
Each token is assigned to a node on the chosen path
Each sentence is assigned to a table
Each table is assigned to a path
Each token is assigned to a node on the chosen path
Each sentence is assigned to a table
Each table is assigned to a path
Each token is assigned to a node on the chosen path
SHLDA: Generating words

Each sentence is assigned to a table

Each table is assigned to a path

Each token is assigned to a node on the chosen path
Each sentence is assigned to a table

Each table is assigned to a path

Each token is assigned to a node on the chosen path
SHLDA: Generating words

- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path
SHLDA: Generating words

- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path
SHLDA: Generating response variable

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]
**SHLDA: Generating response variable**

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]

Diagram:
- Debates
  - Heath Care: D: 0.9, R: 0.0
  - Environment: D: 0.4
  - Economy: R: 0.5
  - Nature: D: 1.4
  - Externalities: R: 0.6
  - Industry: R: 1.5

- Can capture issue-specific polarized words
- Some words are polarized regardless of the issue
- Conservative: freedom, big government, presidential overreach, free market
- Liberal: progressive, fair share, one percent, well regulated

**Agenda-setting & Framing in Political Text**
SHLDA: Generating response variable

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]

Document empirical distribution over nodes

Agenda-setting & Framing in Political Text
SHLDA: Generating response variable

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]

- Can capture issue-specific polarized words

- Some words are polarized regardless of the issue
  - Conservative: freedom, big government, presidential overreach, free market
  - Liberal: progressive, fair share, one percent, well regulated

- Agenda-setting & Framing in Political Text
Can capture issue-specific polarized words

Some words are polarized regardless of the issue
  - Conservative: freedom, big government, presidential overreach, free market
  - Liberal: progressive, fair share, one percent, well regulated
Can capture issue-specific polarized words

Some words are polarized regardless of the issue

- Conservative: freedom, big government, presidential overreach, free market
- Liberal: progressive, fair share, one percent, well regulated
Can capture issue-specific polarized words

Some words are polarized regardless of the issue

- Conservative: freedom, big government, presidential overreach, free market
- Liberal: progressive, fair share, one percent, well regulated
Can capture issue-specific polarized words

Some words are polarized regardless of the issue
  - Conservative: freedom, big government, presidential overreach, free market
  - Liberal: progressive, fair share, one percent, well regulated
## Quantitative results: Regression

### Problem

Predicting response variable for unseen documents

- Predicting ideological leaning for political debates turns
  - Response: DW-NOMINATE score—estimated ideological score of legislators on a liberal-conservative spectrum
- Predicting ratings for product/movie reviews
  - Response: review ratings (1–5 stars)

### Datasets

- 109th congressional floor debates
- Amazon reviews
- Movie reviews

### Evaluation

Mean square error averaged over 5 folds (lower is better)
### Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td>0.860</td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td>1.076</td>
<td>0.871</td>
<td>0.673</td>
</tr>
</tbody>
</table>

Table: Mean squared error averaged over 5 folds (lower is better).
## Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA$_{10}$</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA$_{30}$</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA$_{50}$</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA$_{10}$</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA$_{30}$</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA$_{50}$</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td><strong>0.860</strong></td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA$_{10}$</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA$_{30}$</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA$_{50}$</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td><strong>1.076</strong></td>
<td>0.871</td>
<td><strong>0.673</strong></td>
</tr>
</tbody>
</table>

**Table:** Mean squared error averaged over 5 folds (lower is better).
Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td>0.860</td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td>1.076</td>
<td>0.871</td>
<td>0.673</td>
</tr>
</tbody>
</table>

Table: Mean squared error averaged over 5 folds (lower is better).
Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA$_{10}$</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA$_{30}$</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA$_{50}$</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA$_{10}$</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA$_{30}$</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA$_{50}$</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td>$\textbf{0.860}$</td>
<td>0.702</td>
</tr>
<tr>
<td>SLD$_{A10}$</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLD$_{A30}$</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLD$_{A50}$</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td>$\textbf{1.076}$</td>
<td>0.871</td>
<td>$\textbf{0.673}$</td>
</tr>
</tbody>
</table>

Table: Mean squared error averaged over 5 folds (lower is better).
### Quantitative results: Regression Models

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA_{10}</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA_{30}</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA_{50}</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA_{10}</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA_{30}</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA_{50}</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td><strong>0.860</strong></td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA_{10}</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA_{30}</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA_{50}</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td><strong>1.076</strong></td>
<td>0.871</td>
<td><strong>0.673</strong></td>
</tr>
</tbody>
</table>

**Table:** Mean squared error averaged over 5 folds (lower is better).
## Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td>0.860</td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;10&lt;/sub&gt;</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;30&lt;/sub&gt;</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA&lt;sub&gt;50&lt;/sub&gt;</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td><strong>1.076</strong></td>
<td><strong>0.871</strong></td>
<td><strong>0.673</strong></td>
</tr>
</tbody>
</table>

Table: Mean squared error averaged over 5 folds (lower is better).  

---

**Agenda-setting & Framing in Political Text**
## Quantitative results: Regression

<table>
<thead>
<tr>
<th>Models</th>
<th>Floor Debates</th>
<th>Amazon</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-LDA(_{10})</td>
<td>1.247</td>
<td>1.241</td>
<td>0.970</td>
</tr>
<tr>
<td>SVR-LDA(_{30})</td>
<td>1.183</td>
<td>1.091</td>
<td>0.938</td>
</tr>
<tr>
<td>SVR-LDA(_{50})</td>
<td>1.135</td>
<td>1.130</td>
<td>0.906</td>
</tr>
<tr>
<td>SVR-VOC</td>
<td>1.467</td>
<td>0.972</td>
<td>0.681</td>
</tr>
<tr>
<td>SVR-LDA-VOC</td>
<td>1.101</td>
<td>0.965</td>
<td>0.678</td>
</tr>
<tr>
<td>MLR-LDA(_{10})</td>
<td>1.151</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td>MLR-LDA(_{30})</td>
<td>1.125</td>
<td>1.065</td>
<td>0.936</td>
</tr>
<tr>
<td>MLR-LDA(_{50})</td>
<td>1.081</td>
<td>1.114</td>
<td>0.914</td>
</tr>
<tr>
<td>MLR-VOC</td>
<td>1.124</td>
<td>0.869</td>
<td>0.721</td>
</tr>
<tr>
<td>MLR-LDA-VOC</td>
<td>1.120</td>
<td><strong>0.860</strong></td>
<td>0.702</td>
</tr>
<tr>
<td>SLDA(_{10})</td>
<td>1.145</td>
<td>1.113</td>
<td>0.953</td>
</tr>
<tr>
<td>SLDA(_{30})</td>
<td>1.188</td>
<td>1.146</td>
<td>0.852</td>
</tr>
<tr>
<td>SLDA(_{50})</td>
<td>1.184</td>
<td>1.939</td>
<td>0.772</td>
</tr>
<tr>
<td>SHLDA</td>
<td><strong>1.076</strong></td>
<td>0.871</td>
<td><strong>0.673</strong></td>
</tr>
</tbody>
</table>

Table: Mean squared error averaged over 5 folds (lower is better).
Supervised hierarchical latent Dirichlet allocation (SHLDA):

★ Extends existing supervised topic model using a hierarchy of topics
  - provides a formal computational model corresponding to the theory of framing as second-level agenda setting
★ Combines topic regression with lexical regression to improve predictions
  - improves performance in predicting continuous metadata for unseen documents
In collaboration with Prof. Kristina Miler (Government & Politics, UMD)
In SHLDA, we used pre-computed DW-NOMINATE scores to estimate the positions of legislators on the liberal-conservative spectrum.

One-dimensional ideal points

LIBERAL

CONSERVATIVE
Motivation

- In SHLDA, we used pre-computed DW-NOMINATE scores to estimate the positions of legislators on the liberal-conservative spectrum.
- One-dimensional ideal points

![Diagram showing the spectrum between Liberal and Conservative with differentideal points]

- People might have different positions on different issues.
- → Multi-dimensional ideal points
One-dimensional Ideal Point using Votes

[Poole and Rosenthal, 1985]
Legislator a votes 'Yea' on bill b with probability

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

\[ \Phi(\alpha) = \frac{\exp(\alpha)}{\exp(\alpha) + 1} \]

[Poole and Rosenthal, 1985]
One-dimensional Ideal Point using Votes

**Legislator a votes 'Yea' on bill b with probability**

\[ p(v_{a,b} = \text{Yea}) = \Phi(u_{a}x_{b} + y_{b}) \]

- One-dimensional ideal point of legislator a
- Polarity of bill b
- Popularity of bill b

[Poole and Rosenthal, 1985]
Legislator a votes 'Yea' on bill b with probability

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

One-dimensional ideal point of legislator a

Polarity of bill b

Popularity of bill b

[Poole and Rosenthal, 1985]
Legislator $a$ votes 'Yea' on bill $b$ with probability:

$$p(v_{a,b} = \text{Yea}) = \Phi(u_ax_b + y_b)$$

One-dimensional ideal point of legislator $a$

Polarity of bill $b$

Popularity of bill $b$

[Poole and Rosenthal, 1985]
Legislator a votes 'Yea' on bill b with probability

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

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]
Legislator $a$ votes 'Yea' on bill $b$ with probability

$$p(v_{a,b} = \text{Yea}) = \Phi \left( \sum_{k=1}^{K} u_{a,k} x_{b,k} + y_b \right)$$

Multi-dimensional ideal point of legislator $a$

$K$ dimensions

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]
Legislator a votes 'Yea' on bill b with probability

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

Multi-dimensional ideal point of legislator a

K dimensions

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]
Legislator \( a \) votes 'Yea' on bill \( b \) with probability

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

Multi-dimensional ideal point of legislator \( a \)

Dimensions are difficult to interpret

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]
Multi-dimensional Ideal Point using Votes & Text


Dimensions are difficult to interpret.
Legislator a votes 'Yea' on bill b with probability

\[ p(v_{a,b} = \text{Yea}) = \Phi \left( x_b \sum_{k=1}^{K} u_{a,k} \theta_{b,k} + y_b \right) \]


Dimensions are difficult to interpret
Legislator $a$ votes 'Yea' on bill $b$ with probability

$$p(v_{a,b} = \text{Yea}) = \Phi \left( x_b \sum_{k=1}^{K} u_{a,k} \theta_{b,k} + y_b \right)$$

Multi-dimensional ideal point of legislator $a$

Topic proportion of bill $b$ estimated from its text

Legislator a votes 'Yea' on bill b with probability

\[ p(v_{a,b} = \text{Yea}) = \Phi \left( x_b \sum_{k=1}^{K} u_{a,k} \theta_{b,k} + y_b \right) \]

Multi-dimensional ideal point of legislator a

Topic proportion of bill b estimated from its text

Our approach: Hierarchical Ideal Point Topic Model

Health

- obamacare, patient, doctor, physician, afford_care, hospit, insur, replac, mandat, exchang, health_insur, coverag, medicaid, patient_protect, board

Frame H1

- afford_care, exchang, patient_protect, human_servic, public_health, slush_fund, pppaca, mandator, mandator_spend, governor, hospit, health_center, flexibl, teach_health

Frame H2

- patient, doctor, physician, hospit, medicaid, board, georgia, save_medicar, nurs, tennesse, page, bureaucrat, advisori_board, medicin, independ_payment

Frame H3

- obamacare, replac, mandat, insur, health_insur, coverag, social_secur, premium, repeal_obamacare, entit, govern_takeov, purchas, unconstitut, preexist_condit

Macroeconomics

- balanc_budget, borrow, debt ceil, cap, cut_spend, nation_debt, grandchildren, rais_tax, entit, white_hous, debt_limit, prosper

Frame M1

- white_hous, shut, continu_resolut, mess, hous_republican, novemb, govern_shutdown, senat_reid, harri_reid, vision, shutdown, liber, arriv, republican_parti, blame

Frame M2

- balanc_budget, debt ceil, cap, cut_spend, debt_limit, spend_cut, fiscal_hous, grandchildren, guarante, default, augst, obama, deficit_spend, rein, feder_budget

Frame M3

- borrow, nation_debt, rais_tax, entit, prosper, chart, grandchildren, spend_monei, size, gdp, tax_increas, cent, govern_spend, social_secur
Using both votes and text to learn

- Two-level topic hierarchy
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
- Use existing labeled data to learn priors for interpretable issues
- Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions
Using both votes and text to learn

- **Two-level topic hierarchy**
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
- Use existing labeled data to learn priors for interpretable issues
- Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions
Using both votes and text to learn

- Two-level topic hierarchy
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
  - Use existing labeled data to learn priors for interpretable issues
  - Ideal points for frames for predictions using text only

- Ideal points in multiple interpretable dimensions

Use prior to learn interpretable issue topics
Using both votes and text to learn

- **Two-level topic hierarchy**
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
- Use existing labeled data to learn priors for interpretable issues
- Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions

Learn ideal point for each frame
Approach Overview

Using both votes and text to learn

- **Two-level topic hierarchy**
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
  - Use existing labeled data to learn priors for interpretable issues
  - Ideal points for frames for predictions using text only

- **Ideal points in multiple interpretable dimensions**

Multi-dimensional Ideal Points
Approach Overview

Inputs

- A collection of votes $\{v_{a,b}\}$
- A collection of $D$ speeches $\{w_d\}$, each of which is given by legislator $a_d$
- A collection of $B$ bill text $\{w'_b\}$
Hierarchical Ideal Point Topic Model

Modeling bill text

- Each bill text $b$ is a mixture over $K$ issues $\vartheta_b$
- Each bill token is generated from the topic at a first-level issue node

Health

<table>
<thead>
<tr>
<th>Frame H1</th>
<th>Frame H2</th>
<th>Frame H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>afford_care, exchang, patient_protect, human_servic, public_health, slush_fund, ppace, mandatior, mandatior_spend, governor, hospit, health_center, flexibl, teach_health, obamacar, patient, doctor, physician, afford_care, hospit, insur, replac, mandat, exchang, health insur, coverag, medicaid, patient_protect, board</td>
<td>patient, doctor, physician, hospit, medicaid, board, georgia, save_medrac, nurs, tennesse, page, bureaucrat, advisori_board, medicin, independ_payment, obamacar, replac, mandat, insur, health insur, coverag, social secur, premium, repeal_obamacar, entitt, govern_takeov, purchas, unconstitut, preexist_condit</td>
<td>obamacar, replac, mandat, insur, health insur, coverag, social secur, premium, repeal_obamacar, entitt, govern_takeov, purchas, unconstitut, preexist_condit</td>
</tr>
</tbody>
</table>

Macroeconomics

<table>
<thead>
<tr>
<th>Frame M1</th>
<th>Frame M2</th>
<th>Frame M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>white_hous, shut, continu_resolut, mess, hous_republican, novemb, govern_shutdown, senat_reid, harri_reid, vision, shutdown, liber, arriv, republican_parti, blame</td>
<td>balanc_budget, debt_cell, cap, cut_spend, nation_debt, grandchildren, rais_tax, entitl, white_hous, debt_limit, prosper, nation_debt, balance_budget, cap, cut_spend, debt_limit, prosper, borrom</td>
<td>balanc_budget, debt_cell, cap, cut_spend, debt_limit, spend_cut, fiscal_hous, grandchildren, guarant, default, august, obama, deficit_spend, rein, feder_budget</td>
</tr>
</tbody>
</table>
Hierarchical Ideal Point Topic Model

Modeling speeches

- Each speech $d$ also has a distribution $\theta_d$ over $K$ issues.
- For each issue $k$, each speech $d$ has a distribution over an unbounded number of frames $\psi_{d,k}$.
- Each speech token is generated from the topic at a second-level frame node.
Hierarchical Ideal Point Topic Model

Modeling votes

- Legislator $a$ votes ‘Yea’ on bill $b$ with probability
  \[ p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^{K} \vartheta_{b,k} u_{a,k} + y_b) \]
- Ideal point $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J} \eta_{k,j} \psi_{a,k,j}, \rho)$

Health

- obamacar, patient, doctor, physician, afford_care, hospit, insur, replac, mandat, exchang, health_insur, coverag, medicaid, patient_protect, board

Macroeconomics

- balanc_budget, borrow, debt_cell, cap, cut_spend, nation_debt, grandchildren, rais_tax, entitl, white_hous, debt_limit, prosper

Multi-dimensional Ideal Points from Roll-call Votes and Text
## The Tea Party

- Recent American political movement supporting more freedom, smaller government, lower tax
- Played an important role in recent electoral politics, especially within the Republican Party
- Organizations:
  - Institutional: Tea Party Caucus
  - Other: Tea Party Express, Tea Party Patriots, Freedom Works
- “Conventional views of ideology as a single–dimensional, left-right spectrum experience great difficulty in understanding or explaining the Tea Party.”
  
  [Carmines and D’Amico, 2015, ARPS]
The Tea Party

- Recent American political movement supporting more freedom, smaller government, lower tax
- Played an important role in recent electoral politics, especially within the Republican Party
- Organizations:
  - Institutional: Tea Party Caucus
  - Other: Tea Party Express, Tea Party Patriots, Freedom Works
- “Conventional views of ideology as a single–dimensional, left-right spectrum experience great difficulty in understanding or explaining the Tea Party.”
  [Carmines and D’Amico, 2015, ARPS]

Data

- 240 Republican Representatives in the 112th U.S. House
- 60 are members of the Tea Party Caucus (self-identified)
- 60 key votes selected by Freedom Works (2011-2012)
- Speeches, bill text and voting records from the Library of Congress
One-dimensional Ideal Points

Tea Party Caucus

Multi-dimensional Ideal Points from Roll-call Votes and Text
One-dimensional Ideal Points

Tea Party Caucus

Ideal Point

-1.6 -1.5 -1.4 -1.3 -1.2

Member Nonmember

Judy Biggert
Ander Crenshaw
Jim Gerlach
Michael Grimm
Christopher H. Smith
Harold Rogers
Peter T. King
Patrick Meehan
Rodney M. Alexander

Bob Dold
Steve Stivers
Jon Runyan
Ileana RosLehtinen
Dave G. Reichert
Mario DiazBalart

Multi-dimensional Ideal Points from Roll-call Votes and Text

Trent Franks
Trey Gowdy
Tom McClintock
David Schweikert
Doug Lamborn
Scott Garrett
Joe Walsh
Tom McClinton
David Schweikert
Trey Gowdy
Trent Franks

Ideal Point

1.4 1.6 1.8

Member Nonmember

Jeff Duncan
Justin Amash
Mick Mulvaney
Paul C. Broun
Jim Jordan
Raúl Labrador
Joe Walsh

Jeff Flake
Tom Graves
Raúl Labrador
Joe Walsh

Multi-dimensional Ideal Points from Roll-call Votes and Text
One-dimensional Ideal Points

- **Alexander** and **Crenshaw**'s votes only agree with Freedom Works 48% and 50% respectively.
- Both voted for raising the debt ceiling and are listed as “traitor.”

---

**John T. Reed on Headline News**

<table>
<thead>
<tr>
<th>House Tea Party Caucus members</th>
<th>how they voted on debt ceiling increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sandy Adams, Florida</td>
<td>traitor</td>
</tr>
<tr>
<td>Robert Aderholt, Alabama</td>
<td>traitor</td>
</tr>
<tr>
<td>Todd Akin, Missouri</td>
<td>no</td>
</tr>
<tr>
<td><strong>Rodney Alexander, Louisiana</strong></td>
<td>traitor</td>
</tr>
<tr>
<td>Michele Bachmann, Minnesota, Chairman</td>
<td>no</td>
</tr>
<tr>
<td>Rob Bishop, Utah</td>
<td>no</td>
</tr>
<tr>
<td><strong>Ander Crenshaw, Florida</strong></td>
<td>traitor</td>
</tr>
<tr>
<td>Michael C. Burgess, Texas</td>
<td>traitor</td>
</tr>
</tbody>
</table>

---

Multi-dimensional Ideal Points from Roll-call Votes and Text
Flake and Amash didn’t self-identify as members of the Tea Party Caucus but have been endorsed by other Tea Party organizations.

NEW REPUBLIC

“Some 46 House members and six senators had been listed as part of the loosely organized Tea Party caucus in Congress. In addition, there were about 18 other House members like Trey Gowdy, Mark Meadows, and Justin Amash, and several senators, including Jeff Flake and Pat Toomey, who owed their election to support from the Tea Party and its Washington allies.”
Freedom Works’ key votes on most highly polarized dimensions are about government spending.
## Experiment setup

- **Task**: Binary classification of whether a legislator is a member of the Tea Party Caucus
- **Evaluation metric**: AUC-ROC
- **Classifier**: SVM\textsuperscript{light}
- **Five-fold stratified cross-validation**

**Features**

- **Text-based features**: normalized term frequency (TF) and TF-IDF
- **Vote-based features**: HIPTM features extracted from our model including $K$-dim ideal point $u_a$, $k$ estimated from both votes and text
- **K-dim ideal point estimated from text only**: $\hat{\psi}_a, k$
- **$K$-dim ideal point estimated from votes**: $\sum_{K=1}^{K} \theta b, k u_a, k + y_b$
Experiment setup

- Task: Binary classification of whether a legislator is a member of the Tea Party Caucus
- Evaluation metric: AUC-ROC
- Classifier: SVM\textsuperscript{light}
- Five-fold stratified cross-validation

Features

- Text-based features: normalized term frequency (TF) and TF-IDF
- Vote: binary features
- HIPTM: features extracted from our model including
  - \( K \)-dim ideal point \( u_{a,k} \) estimated from both votes and text
  - \( K \)-dim ideal point estimated from text only \( \eta_k^T \hat{\psi}_{a,k} \)
  - \( B \) probabilities estimating \( a \)'s votes \( \Phi(x_b \sum_{k=1}^{K} \psi_{b,k} u_{a,k} + y_b) \)
Tea Party Caucus Membership Prediction: Votes & Text

AUCROC

0.60
0.65
0.70
0.75

TF TFIDF Vote HIPTM Vote-TF Vote-TF-IDF Vote-HIPTM All

Multi-dimensional Ideal Points from Roll-call Votes and Text
Tea Party Caucus Membership Prediction: Votes & Text

Multi-dimensional Ideal Points from Roll-call Votes and Text
Tea Party Caucus Membership Prediction: Votes & Text

- **AUCROC**
  - 0.60
  - 0.65
  - 0.70
  - 0.75

- **TF**
- **TFIDF**
- **Vote**
- **HIPTM**
- **Vote-TF**
- **Vote-TF-IDF**
- **Vote-HIPTM**
- **All**

**Text-based Features**

**Vote Features**

Multi-dimensional Ideal Points from Roll-call Votes and Text
Tea Party Caucus Membership Prediction: Votes & Text

AUCROC

<table>
<thead>
<tr>
<th>Feature</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFIDF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vote</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIPTM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vote-TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vote-TF-IDF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vote-HIPTM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Text-based Features

Vote Features

Our Features
Tea Party Caucus Membership Prediction: Votes & Text

AUCROC

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>AUCROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-based Features</td>
<td>0.60</td>
</tr>
<tr>
<td>Vote</td>
<td>0.65</td>
</tr>
<tr>
<td>Our Features</td>
<td>0.70</td>
</tr>
<tr>
<td>Combining Vote</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- TF
- TFIDF
- Vote
- HIPTM
- Vote-TF
- Vote-TF-IDF
- Vote-HIPTM
- All

Multi-dimensional Ideal Points from Roll-call Votes and Text
Tea Party Caucus Membership Prediction: Votes & Text

- **AUCROC**
  - 0.60
  - 0.65
  - 0.70
  - 0.75

- **Features**
  - Text-based
    - Vote
    - HIPTM
    - Vote-TF
    - Vote-TF-IDF
    - Vote-HIPTM
  - Our Features
    - Combining Vote with TF/TF-IDF
    - Combining Vote with Our Features

Multi-dimensional Ideal Points from Roll-call Votes and Text
Vote-based features are not needed at test time, so this model makes it possible to do better prediction even for people who have no voting record in Congress e.g., new members of Congress or political candidates.
### AUCROC

<table>
<thead>
<tr>
<th>Model</th>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td>text only</td>
<td>text only</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>text only</td>
<td>text only</td>
</tr>
<tr>
<td>HIPTM</td>
<td>text and votes</td>
<td>text only</td>
</tr>
</tbody>
</table>

Vote-based features are not needed at test time, so this model makes it possible to do better prediction even for people who have no voting record in Congress, e.g., new members of Congress or political candidates.
Vote-based features are not needed at test time, so this model makes it possible to do better prediction even for people who have no voting record in Congress.

- e.g., new members of Congress or political candidates.
We introduce the *Hierarchical Ideal Point Topic Model* which extends existing multi-dimensional ideal point models using a hierarchy of topics, allowing us to

- Discover and analyze **agenda issues and issue-specific frames** in a unified framework
- Provide a formal computational model corresponding to the theory of **framing as second-level agenda setting**
- Analyze ideological positions of legislators in **multiple interpretable dimensions**
- Make **predictions on issue-specific ideological position** of unseen legislators using their **text only**
Summary of Contributions

Technical Contributions

1. Extend prior work on topic segmentation in conversation by incorporating *speaker identity* and using *Bayesian nonparametrics*

2. Capture *dependency among labels* in multi-labeled data using a *tree-structured topic hierarchy*

3. Extend existing supervised topic model using a *hierarchy of topics* and combine topic regression with *lexical regression* to improve prediction

4. Extend existing *multi-dimensional ideal point models* using a *hierarchy of topics*
Summary of Contributions

Applications

1. Study *agendas and agenda control* in political debates and other conversations. Improve performance in topic segmentation and influencer detection.

2. Analyze *policy agenda issues* in legislative text and how they relate to each other using an *interpretable label hierarchy*. Improve performance in predicting held-out words and multiple labels of unseen documents.

3. Study *agenda-setting and framing* in a unified hierarchical framework. Improve performance in ideology prediction and sentiment analysis.

4. Provide a formal computational model corresponding to the theory of *framing as second-level agenda setting*. Analyze ideological positions of legislators in *multiple interpretable*...
### Source code

- All introduced models: [https://github.com/vietansegan](https://github.com/vietansegan)

### Data

- *Crossfire* data:
- Bill text and voting records: available soon (or email me)
Acknowledgment

- Philip Resnik (co-advisor)
- Jordan Boyd-Graber (co-advisor)
- Hal Daumé III
- Héctor Corrada Bravo
- Wayne McIntosh
- Hanna Wallach
- Kristina Miler
- Deborah A. Cai
- Stephen F. Altschul
- Jonathan Chang
- Yuening Hu
- Zhai Ke
- Friends in CLIP
- Funding agencies: NSF#1211153, NSF#1018625, the Army Research Laboratory through ARL Cooperative Agreement W911NF-09-2-0072 and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), the Army Research Laboratory
Thanks!

vietan@cs.umd.edu
www.cs.umd.edu/~vietan
Problem

Study agendas and agenda control in political debates and other conversations

SITS

By modeling explicitly agenda control behaviors of speakers, SITS is able to discover

- the topics discussed in a set of conversations
- how these topics are shared across conversations
- when these topics changes
- a speaker-specific measure of agenda control

Applications

- Analyzing agendas and agenda control behaviors of candidates in political debates (2008 election & 2012 Republican primary debates)
- Improve performance on two quantitative tasks: topic segmentation and influencer detection
Conclusion and Proposed Work
Argviz: Interactive Visualization of Topic Dynamics

Conclusion and Proposed Work
Conclusion and Proposed Work

Argviz: Interactive Visualization of Topic Dynamics

<table>
<thead>
<tr>
<th>Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>44</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>46</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFILL: Governor Palin, is that so?</td>
</tr>
<tr>
<td>PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also. I think that this is important to come back to, with that energy policy plan again that was voted for in '08. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us. We're circulating about $700 billion a year into foreign countries, some who do not like America -- they certainly don't have our best interests at heart -- instead of those dollars circulating here, creating tens of thousands of jobs and allowing domestic supplies of energy to be tapped into and start flowing into these very, very hungry markets. Energy independence is the key to this nation's future, to our economic future, and to our national security. So when we talk about energy plans, it's not just about who got a tax break and who didn't. And we're not giving oil companies tax breaks, but it's about a heck of a lot more than that. Energy independence is the key to America's future.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>bridge street economy</td>
</tr>
<tr>
<td>mccain people wall washington home</td>
</tr>
<tr>
<td>campaign people running talking american campaigns tough congressman</td>
</tr>
<tr>
<td>trade free agreement country south priorities families agreements</td>
</tr>
<tr>
<td>genocide security darfur tell bosnia marines national force</td>
</tr>
<tr>
<td>question tonight thank senator presidential debate mccain obama</td>
</tr>
</tbody>
</table>

Visualizations and computational models are used to identify topic dynamics and shifts in the debate. The visualizations help in understanding the progression of topics and the alignment of speakers with their agendas.
Argviz: Interactive Visualization of Topic Dynamics

Conclusion and Proposed Work

Number two, with regard to bankruptcy now, Gwen, what we should be doing now -- and Barack Obama and I support it -- we should be allowing bankruptcy courts to be able to re-adjust not just the interest rate you're paying on your mortgage to be able to stay in your home, but be able to adjust the principal that you owe, the principal that you owe. That would keep people in their homes, actually help banks by keeping it from going under. But John McCain, as I understand it -- I'm not sure of this, but I believe John McCain and the governor don't support that. There are ways to help people now. And there -- ways that we're offering are not being supported by -- by the Bush administration nor do I believe by John McCain and Governor Palin.

PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also. I think that this is important to come back to, with that energy policy plan again that was voted for in '06. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us. We're circulating about $700 billion a year into foreign countries, some who do not like America -- they certainly don't have our best interests at heart -- instead of those dollars circulating here, creating tens of thousands of jobs and allowing domestic supplies of energy to be tapped into and start flowing into these very, very hungry markets. Energy independence is the key to this nation's future, to our economic future, and to our national security. So when we talk about energy plans, it's not just about who got a tax break and who didn't. And we're not giving oil companies tax breaks, but it's about a heck of a lot more than that. Energy independence is the key to America's future.

IFILL: Governor, I'm happy to talk to you in this next section about energy issues. Let's talk about climate change. What is true and what is false about what we have heard, read, discussed, debated about the causes of climate change?
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden

Well, mortgage-holders didn't pay the price. Only 10 percent of the people who are – have been affected by this whole switch from Chapter 7 to Chapter 13 – it gets complicated. But the point of this – Barack Obama saw the glass as half-empty. I saw it as half-full. We disagreed on that, and 85 senators voted one way, and 15 voted the other way. But here’s the deal. Barack Obama pointed out two years ago that there was a subprime mortgage . . . And there – ways that we’re offering are not being supported by – by the Bush administration nor do I believe by John McCain and Governor Palin.
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden

Well, mortgage-holders didn’t pay the price. [...]

Sarah Palin

That is not so, but because that’s just a quick answer, I want to talk about, again, my record on energy ... When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent ... East Coast politicians who don’t allow energy-producing states like Alaska to produce these, to tap into them, and instead we’re relying on foreign countries to produce for us.
Nonparametric Bayesian topic model for agenda control
Nonparametric Bayesian topic model for agenda control

- Each turn: a multinomial distribution over topics
Nonparametric Bayesian topic model for agenda control

- Each turn: a **multinomial distribution over topics**
- Each speaker: a **biased coin** capturing how likely the speaker changes topic
Nonparametric Bayesian topic model for agenda control

- Each turn: a **multinomial distribution over topics**
- Each speaker: a **biased coin** capturing how likely the speaker changes topic
- Each turn: a **binary latent variable** indicating whether the topic is shifted
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Generative Process

Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden

Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage...
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden

Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage ...
Gwen Ifill

Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden

Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage ...
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage...
Gwen Ifill
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Joe Biden
Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage...
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Well, mortgage-holders didn’t pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .

That is not so, but because that’s just a quick answer, I want to talk about, again, my record on energy . . .
Posterior Inference

Posterior inference task

Given the observed conversational data, the goal is to infer
- the topic distribution of each conversational turn
- when the topic of the conversation changes
- how likely each speaker changes the topic of the conversation

Gibbs sampling

Alternates between
- **Sampling topic assignments**: which topic each token belongs to
- **Sampling topic shift indicator**: if topic shift occurs in each turn
In presidential debates, moderators have much higher scores than candidates do. In the VP debate, IFILL's score is only slightly higher than those of PALIN and BIDEN.

Conclusion and Proposed Work
In presidential debates, moderators have much higher scores than candidates do. In the VP debate, IFILL’s score is only slightly higher than those of PALIN and BIDEN.

Conclusion and Proposed Work
In presidential debates, moderators have much higher scores than candidates do.
In presidential debates, moderators have much higher scores than candidates do.

In the VP debate, IFILL’s score is only slightly higher than those of PALIN and BIDEN.
In presidential debates, moderators have much higher scores than candidates do.

In the VP debate, IFILL’s score is only slightly higher than those of PALIN and BIDEN.

Ifill's questioning and moderating was, as The Atlantic's James Fallows remarked, "terrible." She asked open-ended, utterly predictable questions which presented very little challenge to the candidates. But even more important to the McCain campaign's strategy, Palin was able to simply ignore the questions and recite her talking points.
Conclusion and Proposed Work

Argviz: Interactive Visualization of Topic Dynamics

Number two, with regard to bankruptcy now, Gwen, what we should be doing now -- and Barack Obama and I support it -- we should be allowing bankruptcy courts to be able to re-adjust not just the interest rate you're paying on your mortgage to be able to stay in your home, but be able to adjust the principal that you owe, the principal that you owe. That would keep people in their homes, actually help banks by keeping it from going under. But John McCain, as I understand it -- I'm not sure of this, but I believe John McCain and the governor don't support that. There are ways to help people now. And there -- ways that we're offering are not being supported by -- by the Bush administration nor do I believe by John McCain and Governor Palin.

IFILL: Governor Palin, is that so?

PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also. I think that this is important to come back to, with that energy policy plan again that was voted for in '05. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us. We're circulating about $700 billion a year into foreign countries, some who do not like America -- they certainly don't have our best interests at heart -- instead of those dollars circulating here, creating tens of thousands of jobs and allowing domestic supplies of energy to be tapped into and start flowing into these very, very hungry markets. Energy independence is the key to this nation's future, to our economic future, and to our national security. So when we talk about energy plans, it's not just about who got a tax break and who didn't. And we're not giving oil companies tax breaks, but it's about a heck of a lot more than that. Energy independence is the key to America's future.

IFILL: Governor, I'm happy to talk to you in this next section about energy issues. Let's talk about climate change. What is true and what is false about what we have heard, read, discussed, debated about the causes of climate change?
Conclusion and Proposed Work

Argviz: Interactive Visualization of Topic Dynamics

Computational model identifies turn 48 in the debate as clearly indicating topic control.

Computational model identifies what is on Palin's agenda in this topic shift.

Number two, with regard to bankruptcy now, Gwen, what we should be doing now -- and Barack Obama and I support it -- we should be allowing bankruptcy courts to be able to re-adjust not just the interest rate you're paying on your mortgage to be able to stay in your home, but be able to adjust the principal that you owe, the principal that you owe. That would keep people in their homes, actually help banks by keeping it from going under. But John McCain, as I understand it -- I'm not sure of this, but I believe John McCain and the governor don't support that. There are ways to help people now. And there -- ways that we're offering are not being supported by -- by the Bush administration nor do I believe by John McCain and Governor Palin.

IFILL: Governor Palin, is that so?

PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also. I think that this is important to come back to, with that energy policy plan again that was voted for in '05. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us. We're circulating about $700 billion a year into foreign countries, some who do not like America -- they certainly don't have our best interests at heart -- instead of those dollars circulating here, creating tens of thousands of jobs and allowing domestic supplies of energy to be tapped into and start flowing into these very, very hungry markets. Energy independence is the key to this nation's future, to our economic future, and to our national security. So when we talk about energy plans, it's not just about who got a tax break and who didn't. And we're not giving oil companies tax breaks, but it's about a heck of a lot more than that. Energy independence is the key to America's future.

IFILL: Governor, I'm happy to talk to you in this next section about energy issues. Let's talk about climate change. What is true and what is false about what we have heard, read, discussed, debated about the causes of climate change?
Argviz: Interactive Visualization of Topic Dynamics

Conclusion and Proposed Work
Argviz: Interactive Visualization of Topic Dynamics

Conclusion and Proposed Work
Quantitative Evaluations

- Topic segmentation
- Influencer detection
Task

Divide conversation into smaller, topically coherent segments
Topic Segmentation

Task
Divide conversation into smaller, topically coherent segments

Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Speakers</th>
<th>Conversations</th>
<th>Annotations</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICSI Meetings</td>
<td>60</td>
<td>75</td>
<td>segmentation</td>
<td>engineering</td>
</tr>
<tr>
<td>2008 Debates</td>
<td>9</td>
<td>4</td>
<td>segmentation</td>
<td>politics</td>
</tr>
</tbody>
</table>
Task
Divide conversation into smaller, topically coherent segments

Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Speakers</th>
<th>Conversations</th>
<th>Annotations</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICSI Meetings</td>
<td>60</td>
<td>75</td>
<td>segmentation</td>
<td>engineering</td>
</tr>
<tr>
<td>2008 Debates</td>
<td>9</td>
<td>4</td>
<td>segmentation</td>
<td>politics</td>
</tr>
</tbody>
</table>

Evaluation metrics: WindowDiff

- sliding windows of size $k$ through the conversation
- penalize the window in which the number of boundaries in the model’s segmentation is different from that in the true segmentation
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.
Consider sliding windows of size $k$ and penalize the window in which the numbers of boundaries in the true segmentation and in the model’s segmentation are different.

<table>
<thead>
<tr>
<th>Conversation Turns</th>
<th>True Segmentation</th>
<th>Model's Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Conversation Turns" /></td>
<td><img src="image2" alt="True Segmentation" /></td>
<td><img src="image3" alt="Model's Segmentation" /></td>
</tr>
</tbody>
</table>
Segmentation Performance

- **TextTiling** [?] -
- **P-NoSpeaker-S and P-NoSpeaker-M**:
  parametric, no speaker identity [?]
- **NP-HMM**:
  nonparametric, no speaker identity, single topic per turn [?]
- **P-SITS and NP-SITS**:
  parametric and nonparametric SITS
Influencer detection:

- Detecting speakers who have persuasive abilities over where the conversation is headed and what topics are covered
- Focus of much research in communication, sociology and psychology for decades
Influencer Detection

- Influencer detection:
  - Detecting speakers who have persuasive abilities over where the conversation is headed and what topics are covered
  - Focus of much research in communication, sociology and psychology for decades
- Topic control and management is one of the most effective ways

“the ability to change topical focus, especially given strong cultural and social pressure to be relevant, means having enough interpersonal power to take charge of the agenda”

[Palmer 1989]
Influencer Detection

- Influencer detection:
  - Detecting speakers who have persuasive abilities over where the conversation is headed and what topics are covered
  - Focus of much research in communication, sociology and psychology for decades
  - Topic control and management is one of the most effective ways

“the ability to change topical focus, especially given strong cultural and social pressure to be relevant, means having enough interpersonal power to take charge of the agenda”  
[Palmer 1989]

Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Speakers</th>
<th>Conversations</th>
<th>Annotations</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossfire</td>
<td>2567</td>
<td>1134</td>
<td>influencer</td>
<td>politics</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>604</td>
<td>1991</td>
<td>influencer</td>
<td>varied</td>
</tr>
</tbody>
</table>
## Influencer Detection

### Conclusion and Proposed Work

<table>
<thead>
<tr>
<th>Measure</th>
<th>Crossfire</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted topic shifts</td>
<td>F1</td>
<td>AUC.ROC</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td>Total topic shifts</td>
<td>0.75</td>
<td>0.70</td>
</tr>
<tr>
<td>Total turn lengths</td>
<td>0.65</td>
<td>0.60</td>
</tr>
<tr>
<td>Num. of turns</td>
<td>0.50</td>
<td>0.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Crossfire</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted topic shifts</td>
<td>0.70</td>
<td>0.65</td>
</tr>
<tr>
<td>Total topic shifts</td>
<td>0.65</td>
<td>0.60</td>
</tr>
<tr>
<td>Total turn lengths</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>Num. of turns</td>
<td>0.45</td>
<td>0.40</td>
</tr>
</tbody>
</table>
## Influencer Detection

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>AUC.ROC</th>
<th>AUC.PRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted topic shifts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total topic shifts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total turn lengths</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num. of turns</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Crossfire</th>
<th>Wikipedia</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted topic shifts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total topic shifts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total turn lengths</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num. of turns</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduce a nonparametric Bayesian model to discover
- the topics used in a set of conversations
- when these topics change during conversations
- a speaker-specific measure of “agenda control”
Introduce a nonparametric Bayesian model to discover
- the topics used in a set of conversations
- when these topics change during conversations
- a speaker-specific measure of “agenda control”

The model:
- requires low cost: data-driven using texts with available meta-data (i.e, speaker identity)
- provides insights about agenda control in political debates
- improves performances in two computational tasks: topic segmentation and influencer detection
Initialization

Initialize the label tree using the maximum spanning tree on $G$ (Chu-Liu/Edmonds’ algorithm)

MCMC Inference

Alternating between

1. Sampling the node assignment for each token
2. Sampling the topic $\phi$ at each node
3. Updating the tree structure by
   - Proposing a new parent node for each node
   - Accepting/Rejecting the proposal using Metropolis-Hastings algorithm
Posterior Inference

Initialization

Initialize the label tree using the maximum spanning tree on $G$ (Chu-Liu/Edmonds’ algorithm)

MCMC Inference

Alternating between

1. Sampling the node assignment for each token
2. Sampling the topic $\phi$ at each node
3. Updating the tree structure by
   - Proposing a new parent node for each node
   - Accepting/Rejecting the proposal using Metropolis-Hastings algorithm
**Held-out Word Prediction**

<table>
<thead>
<tr>
<th></th>
<th>109</th>
<th>110</th>
<th>111</th>
<th>112</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perplexity</td>
<td>300</td>
<td>250</td>
<td>200</td>
<td>150</td>
</tr>
</tbody>
</table>

LDA
Unsupervised topic model, no label constraint

L-LDA
Labeled-LDA (Ramage et al '09) one topic per label, flat structure, *hard assignment* (i.e., document can only be generated from its fixed set of topics)

L2F
Label-to-Flat hierarchy simplified version of L2H with a fixed flat hierarchy; allows *soft assignments* in contrast to Labeled-LDA

L2H
Label-to-Hierarchy
Our proposed model

---

- soft assignments and hard assignments perform similarly
- using the hierarchy improves significantly
- with no label constraint, LDA achieves best perplexity

**Conclusion and Proposed Work**
Environment

**Envtl. assessment, monitoring, research**
- oil, outer_continent, oil_spill, shelf_land, oil_pollut, dispers, royalti, coastal, shelf, coast_guard, respons_plan, hazard_substanc, chemic

**Marine and coastal resources, fisheries**
- gulf_coast, gulf, coastal, fisheri, marin, ocean, ecosystem, marin_debri, fisheri_conserv, atmospher_administr, trust_fund, nation_ocean

**Wildlife conservation and habitat protection**
- coral_reef, speci, ecosystem, endang_speci, nation_wildlif, salmon_stronghold, anim, salmon, livestock, joint_ventur, laci_act, wildlf_foundat, refug

**Water quality**
- chesapeake_bai, watersh, basin, feder_water, pollut_control, water_qualiti, island_sound, restor_activ, sediment, lake_taho, water_pollut, river

**Water use and supply**
- navajo_nation, hopi_tribe, settlement_agreement, river, restor_agreement, lower_colorado, river_water, colorado_river, water_qualiti

**Environmental protection**
- green_infrastructur, hypoxia, ballast_water, harm_algal, bloom, commerci_vessel, vessel, mercuri, lake, pollut_control, marin, speci

**Environmental regulatory procedures**
- chemic_substanc, substanc, chemic, safeti_standard, cover_water, processor, cover_treatment, mixtur, administr_determin, intent_act

**Oil and gas**
- oil, outer_continent, leas_sale, coastal_plain, shelf_land, leas_program, pipelin, shelf, drill, coastal, feder_land, lesse, polici_act, gulf, keystone, royalti

**Conclusion and Proposed Work**
Agenda setting & Framing

**Agenda setting**

- The salient issues are considered important by the public
- **What topics are talked about?**
- Example: 0.9 correlation between what people thought was the most important election issues and what the local and national media reported was the most important issues
## Agenda setting & Framing

### Agenda setting
- The salient issues are considered important by the public
- **What** *topics are talked about?*
- Example: 0.9 correlation between *what people thought* was the most important election issues and *what the local and national media reported* was the most important issues

### Framing
- The way an issue is presented influences or encourages particular perspectives or interpretations
- **How** *are the topics talked about?*
- Example: Story on *marijuana* emphasizes the *cost of drug war* and the potential for revenue through legalizing/regulation of the market → *economic frame*
# Agenda setting & Framing

## Agenda setting
- The salient issues are considered important by the public
- **What** topics are talked about?

## Framing
- The way an issue is presented influences or encourages particular perspectives or interpretations
- **How** are the topics talked about?

---

"It’s not what you say, it’s **how** you say it"

Frank Luntz (1997)
Republican Party strategist

"Don’t Think of an Elephant!: Know Your Values and **Frame** the Debate"

George Lakoff (2004)
Democratic Party advisor
Agenda setting & Framing

Framing: second-level agenda setting

- Agenda setting: which issues are salient
- Framing: which aspects of the discussed issues are salient
Framing: second-level agenda setting

- Agenda setting: which *issues* are salient
- Framing: which *aspects* of the discussed issues are salient
Agenda setting & Framing

**Framing**: second-level agenda setting

- Agenda setting: which *issues* are salient
- **Framing**: which *aspects* of the discussed issues are salient

![Diagram]

- Debates
  - Heath Care
  - Environment
  - Economy
  - Nature
  - Externalities
  - Industry
1. For each node $k \in [1, \infty)$ in the tree
   (a) Draw topic $\phi_k \sim \text{Dir}(\beta_k)$
   (b) Draw regression parameter $\eta_k \sim \mathcal{N}(\mu, \sigma)$
2. For each word $v \in [1, V]$, draw $\tau_v \sim \text{Laplace}(0, \omega)$
3. For each document $d \in [1, D]$
   (a) Draw level distribution $\theta_d \sim \text{GEM}(m, \pi)$
   (b) Draw table distribution $\psi_d \sim \text{GEM}(\alpha)$
   (c) For each table $t \in [1, \infty)$, draw a path $c_{d,t} \sim \text{nCRP}(\gamma)$
   (d) For each sentence $s \in [1, S_d]$, draw a table indicator $t_{d,s} \sim \text{Mult}(\psi_d)$
      i. For each token $n \in [1, N_{d,s}]$
         A. Draw level $z_{d,s,n} \sim \text{Mult}(\theta_d)$
         B. Draw word $w_{d,s,n} \sim \text{Mult}(\phi_{c_{d,t_{d,s}},z_{d,s,n}})$
   (e) Draw response $y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$:
      i. $\bar{z}_{d,k} = \frac{1}{N_{d,s}} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} I[k_{d,s,n} = k]$
      ii. $\bar{w}_{d,v} = \frac{1}{N_{d,s}} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} I[w_{d,s,n} = v]$
Data

- A collection of documents $w_1, w_2, \cdots, w_D$
- Each document $d$ has an associated response variable $y_d$
Data

- A collection of documents $w_1, w_2, \ldots, w_D$
- Each document $d$ has an associated response variable $y_d$

Political debates

- $w_d$: debate turn $=$ document

ObamaCare fight reenergizes tea party movement

- $y_d$: ideology of speaker

Democrat

Republican

-1

+1
Data

- A collection of documents $w_1, w_2, \cdots, w_D$
- Each document $d$ has an associated response variable $y_d$

Political debates

- $w_d$: debate turn = document

Obama Blames Boehner for ‘Reckless Republican Shutdown’

- $y_d$: ideology of speaker

Democratic | Republican
---|---
-1 | +1
Qualitative results

Conclusion and Proposed Work

A: minimum_wage
commission
independent_commission
investigate
hurricane_katrina
increase investigation
response fema

D: 1.7

B: flag
constitution
freedom
supreme_court
elections
rights
continuity
american_flag
constitutional_amendment
symbol

R: 1.1

C: billion
budget
children
cuts
debt
tax_cuts
programs
child_support
deficit
education
students
health_care
republicans
national_debt
veterans

D: 4.5

D: bill
speaker
time
amendment
chairman
people
gentleman
legislation
congress
support
country
house
vote

R: 0

E: percent
tax
economy
estate_tax
capital_gains
money
taxes
businesses
families
tax_cuts
pay
tax_relief
social_security
million
alternative_minimum

R: 0.4

R: 1.0

Gses
credit_rating
fannie_mae
regulator
freddie_mac
market
financial_services
agencies
competition
investors
fannie
sec
nationally_recognized

REPUBLICAN

DE DEMOCRAT

D: 2.2

affordable_housing
housing
manager
fund
activities
funds
organizations
voter_registration
faithbased
restrictions
participate
nonprofits

D: 4.3

death_tax
jobs
businesses
business
family_businesses
equipment
repeal_permanency
employees
capital
farms
productivity
sites
freedom

R: 2.2

D: 4.3

affordable_housing
housing
manager
fund
activities
funds
organizations
voter_registration
faithbased
restrictions
participate
nonprofits

D: 4.3
Qualitative results

transmitter ipod car frequency iriver product transmitters live station presets itrip iriver_aft charges international_mode driving

tried waste batteries tunecast rabbit_ears weak terrible antenna hear returned refund returning item junk return

months loves hammock splash love baby drain eurobath hot fits wash play infant secure slip

tub baby water bath sling son daughter sit bathtub sink newborn months bath tub bathe bottom

tivo adapter series adapters phone_line tivo_wireless transfer plugged wireless_adapter tivos plug dvr tivo_series tivo_box tivo_unit

leaks leaked leak leaking hard waste snap suction_cups lock tabs difficult bottom tub_leaks properly ring

router firmware ddwrt wrt54gl version wrt54g tomato linksys linux routers flash versions browser dink stable

version hours phone firmware told spent linksys tech_support technical_support netgear tried customer_service range_expander support return

leak formula bottles_leak feeding leaked brown frustrating started clothes waste newborn playtex_ventaire soaked matter switched

appointments organized phone lists handheld organizer photos etc pictures memos track bells books purse whistles

months noise_cancelling noise son exposed noise_cancellation stopped wires warranty noise_cancelling_bud pay white_noise disappointed

comfortable sound phones sennheiser bass px100 px100s phone headset highs portapros portapro price wear koss

 Conclusion and Proposed Work

POSITIVE

NEGATIVE
Qualitative results

router setup network expander set signal wireless connect linksys connection house wireless_router laptop computer wre54g

monitor radio weather_radio night baby range alerts sound sony house interference channels receiver static alarm

tivo adapter series adapters phone_line tivo_wireless transfer plugged wireless_adapter tivos plug dvr tivo_series tivo_box tivo_unit

hear feature static monitors set live warning volume counties noise outside alert breathing rechargeable_battery alerts

router firmware ddwrt wrt54gl version wrt54g tomato linksys linux routers flash versions browser dlink stable

version hours phone firmware told spent linksys tech_support technical_support netgear tried customer_service range_expander support return
\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]
$y_d \sim \mathcal{N}(\eta^T \tilde{z}_d, \rho)$
**SHLDA: Generating response variable**

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]

### Document empiric distribution over nodes

- **Health Care**: D: 0.9, R: 0.0
- **Environment**: D: 0.4, R: 0.5
- **Nature**: D: 1.4
- **Externalities**: R: 0.6
- **Industry**: R: 1.5

### Context-specific contributions (topics)
- "unpredictable": good for books, bad for steering
- "wonderful", "awesome": always good
- "horrible", "awful": always bad

---

**Conclusion and Proposed Work**
Some words have **context-specific** contributions (topics)
- “unpredictable”: good for books, bad for steering
Some words have **context-specific** contributions (topics)
- “unpredictable”: good for books, bad for steering

Some words have **constant** contributions (words)
- “wonderful”, “awesome”: always good
- “horrible”, “awful”: always bad

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho) \]
Some words have **context-specific** contributions (topics)
- “unpredictable”: good for books, bad for steering

Some words have **constant** contributions (words)
- “wonderful”, “awesome”: always good
- “horrible”, “awful”: always bad

\[
y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)
\]
Some words have **context-specific** contributions (topics)
- “unpredictable”: good for books, bad for steering

Some words have **constant** contributions (words)
- “wonderful”, “awesome”: always good
- “horrible”, “awful”: always bad

\[ y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho) \]
Some words have **context-specific** contributions (topics)
- “unpredictable”: good for books, bad for steering

Some words have **constant** contributions (words)
- “wonderful”, “awesome”: always good
- “horrible”, “awful”: always bad

\[
y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)
\]
Agenda setting & Framing

The swine flu finding only fuels fears from law enforcement along the border who say the illegal immigrants are not being properly screened for diseases and contagious sicknesses before moving along to other facilities for holding across the nation.

“Some of the children who have come to this country may not have a valid legal basis to remain, but some will. Yet, it is virtually impossible for a child to assert a valid claim under immigration law in the absence of legal representation. ... It is a fantasy to believe that unrepresented children have a fair shot in an immigration proceeding”

Issue-specific Framing: Health

Health
- obamacare, patient, doctor, physician, afford_care, hospit, insur, replac, mandat, exchang, health_insur, coverag, medicaid, patient_protect, board

Frame H1
- afford_care, exchang, patient_protect, human_servic, public_health, slush_fund, ppaca, mandatori, mandatori_spend, governor, hospit, health_center, flexibl, teach_health, unlimit

Frame H2
- patient, doctor, physician, hospit, medicaid, board, georgia, save_medicar, nurs, tennesse, page, bureaucrat, advisori_board, medicin, independ_payment

Frame H3
- obamacare, replac, mandat, insur, health_insur, coverag, social_secur, premium, repeal_obamacar, entitl, govern_takeov, purchas, unconstit, preexist_condit, employ

Conclusion and Proposed Work
Issue-specific Framing: Labor & Employment

Labor, Employment, and Immigration

employ, hire, job_creator, union, south_carolina, busi_owner, nlrb, uncertainti, boe, labor, mandat, capit, manufactur, econom_growth

Frame L1

hire, job_creator, busi_owner, uncertainti, employ, capit, manufactur, econom_growth, innov, mandat, emploi, certaini, entrepreneur, regulatori, arkansa

Frame L2

union, south_carolina, nlrb, boe, employ, labor, contractor, wage, locat, board, execut, nation_labor, relat_board, unemploy_rate, project_labor

Conclusion and Proposed Work
**Political agendas.**

**Ideology and interests in the political marketplace.**

**The new look in political ideology research.**
*Annual Review of Political Science*, 18(4).

**The statistical analysis of roll call data.**

**Framing: Toward clarification of a fractured paradigm.**

**How they vote: Issue-adjusted models of legislative behavior.**

**Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators.**
Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking.  

Scaling politically meaningful dimensions using texts and votes.  

*Setting the agenda: The mass media and public opinion*.  
John Wiley & Sons.

A spatial model for legislative roll call analysis.  

How to analyze political attention with minimal assumptions and costs.  

The utility of text: The case of Amicus briefs and the Supreme Court.  
In *Association for the Advancement of Artificial Intelligence*.  