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

Outline

Agenda-setting

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]

Agenda-setting

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]

Energy

Health

Defense

Education

Economics

Immigration

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?

Framing

"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]

Framing

"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]

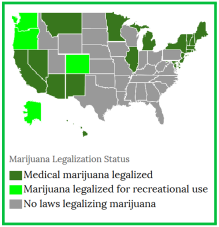
"To frame is to *select some aspects of a perceived reality and make them more salient in a communicating text*, *in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation* for the item described"

[Entman, 1993]

Legalizing Marijuana

Economic frame

Health frame



Legal frame

Framing

"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

Approach	**Costs** [Quinn et al., 2010]		
	Pre-analysis	**Analysis**	**Post-analysis**
Manual content analysis approach			
Close reading			
Human coding			
Automated content analysis approach			
Supervised learning			
Topic modeling			

Content Analysis Approaches

Approach	**Costs** [Quinn et al., 2010]		
	Pre-analysis	**Analysis**	**Post-analysis**
Manual content analysis approach			
Close reading			
Human coding			
Automated content analysis approach			
Supervised learning			
Topic modeling			

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

Approach	Costs [Quinn et al., 2010]		
	Pre-analysis	**Analysis**	**Post-analysis**
Manual content analysis approach			
Close reading	low	high	high
Human coding	high	high	low
Automated content analysis approach			
Supervised learning			
Topic modeling			

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

Approach	Costs [Quinn et al., 2010]		
	Pre-analysis	**Analysis**	**Post-analysis**
Manual content analysis approach			
Close reading	low	high	high
Human coding	high	high	low
Automated content analysis approach			
Supervised learning	high	low	low
Topic modeling	low	low	moderate

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
immigration; illegal immigration; border patrol; border security; agent; alien; illegal alien; deport; southern border; visa; citizenship	tax relief; revenue; tax cut; economic growth; trillion; raising tax; tax increase; tax policiy; cut tax; american family; fiscal; tax revenue	cell; embryo; patient; stem cell; disease; embryonic stem; doctor; physician; medicine; cure; nih; adult stem; stage; drug; ethic	agriculture; animal; farmer; usda; horse; label; manufacture; food safety; meat; rancher; farm; eat; plant; livestock; slaughter	oil; coal; drill; gasoline; ethanol; electric; gallon; car; peak; pump; plant; burn; crude oil; shelf; gulf; refinery

Figure: Typical output of unsupervised topic models

Content Analysis Approaches

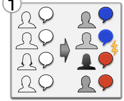
Approach	**Costs** [Quinn et al., 2010]		
	Pre-analysis	**Analysis**	**Post-analysis**
Manual content analysis approach			
Close reading	low	high	high
Human coding	high	high	low
Automated content analysis approach			
Supervised learning	high	low	low
Topic modeling	low	low	moderate

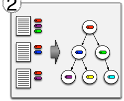
In this thesis

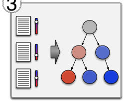
Following the topic modeling approach, we develop a series of new models, which are **guided by additional information associated with the text** and **designed to discover and analyze agenda-setting and framing at a lower cost**.

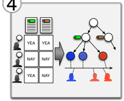
Main Contributions

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

1


2


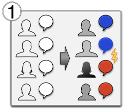
3


4


Technical Contributions

Applications

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)

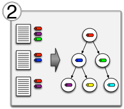
Applications

- ⋆ 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

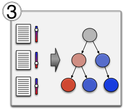
- ⋆ Analyze policy agenda issues in legislative text and how they relate to each other
- ⋆ Learn an interpretable label hierarchy to reduce post-analysis cost
- ⋆ Improve performance in predicting held-out words and multiple labels of unseen documents

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

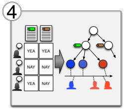
Goal: Study agenda-setting and framing at a lower cost

Technical Contributions

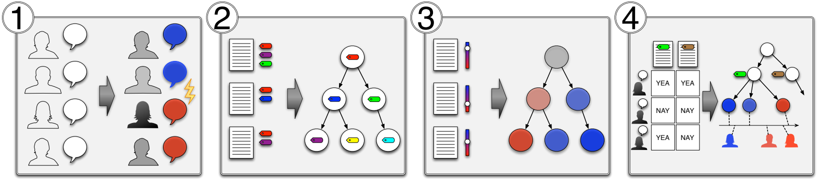
- ⋆ Extend existing **multi-dimensional ideal point** models using a **hierarchy of topics**
- ⋆ *Hierarchical Ideal Point Topic Model* (HIPTM)

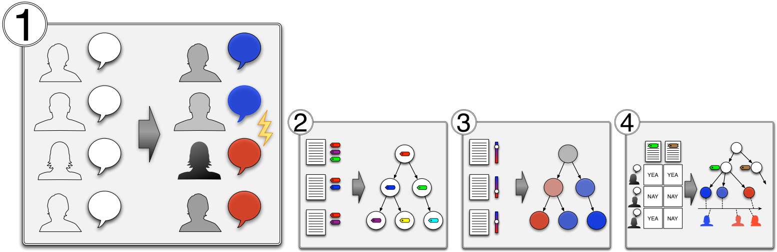
Applications

- ⋆ 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

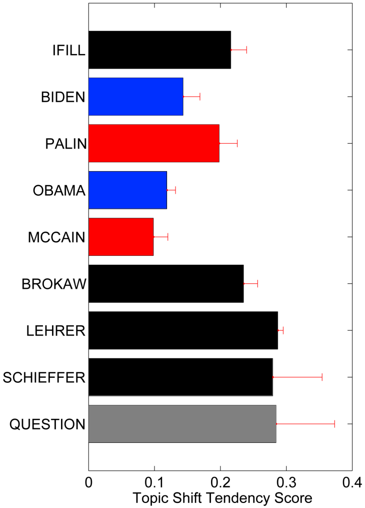


Outline





- **V.-A Nguyen**, J. Boyd-Graber, P. Resnik. *SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations* (**ACL**, 2012)
- **V.-A Nguyen**, Y. Hu, J. Boyd-Graber, P. Resnik. *Argviz: Interactive Visualization of Topic Dynamics in Multi-party Conversations* (**NAACL**, 2013)
- **V.-A Nguyen**, J. Boyd-Graber, P. Resnik, D. Cai, J. Midberry, Y. Wang. *Modeling Topic Control to Detect Influence in Conversations using Nonparametric Topic Models* (**Machine Learning Journal**, 2014)



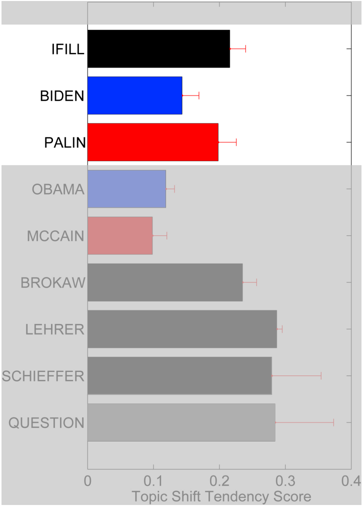
- In presidential debates, moderators have much higher scores than candidates do
- In the VP debate, IFILL's score is only slightly higher than those of PALIN and BIDEN

The Ifill Factor

By Scott Horton

HARPER'S MAGAZINE

…. 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.



- 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

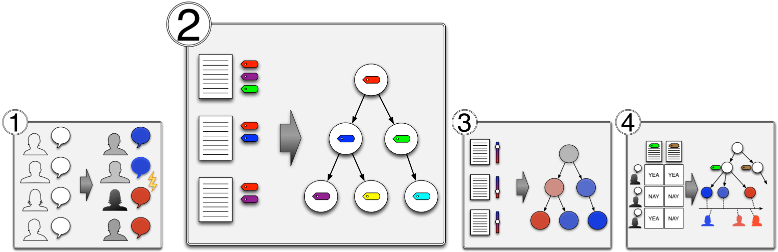
The Ifill Factor

By Scott Horton

HARPER'S MAGAZINE

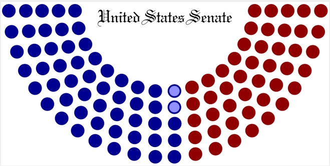
…. Ifill's questioning and moderating was, as The Atlantic's James Fallows remarked, "terrible." She asked open-ended, utterly predictable questions which presented very little challenge to the candidates. But even more important to the McCain campaign's strategy, Palin was able to simply ignore the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

Political Agenda in Legislative Texts

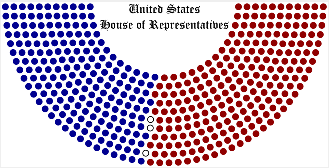


- ⋆ **V.-A. Nguyen**, J. Boyd-Graber, J. Chang, P. Resnik. *Tree-based Label Dependency Topic Models* (**NIPS** Workshop on Topic Models, 2013)
- ⋆ **V.-A. Nguyen**, J. Boyd-Graber, P. Resnik, J. Chang. *Learning a Concept Hierarchy from Multi-labeled Documents* (**NIPS**, 2014)

Senate



House of Representatives



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

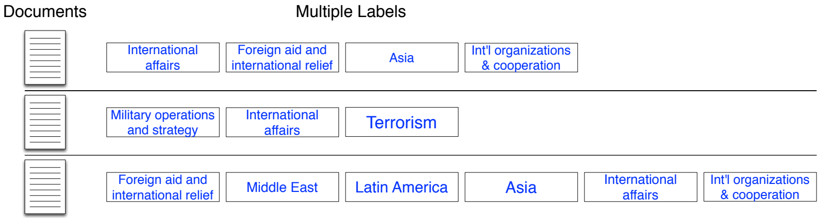
- Define codebook, train coders, annotate documents
  - Policy Agendas Project: define 19 major topics, 225 subtopics
  - Congressional Bills Project: one major topic for each bill
- ↓ Many bills are about more than one topic
- ↓ Difficult to extend over time and to other domains, e.g.,
  - "Immigration" 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

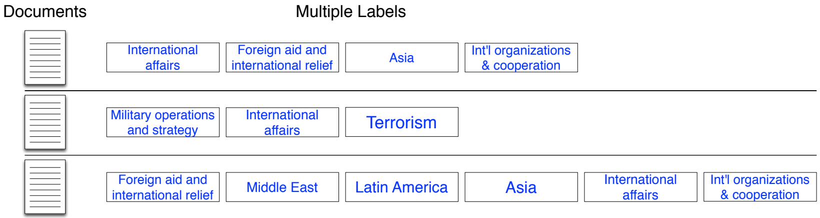
Multi-labeled Data

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



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

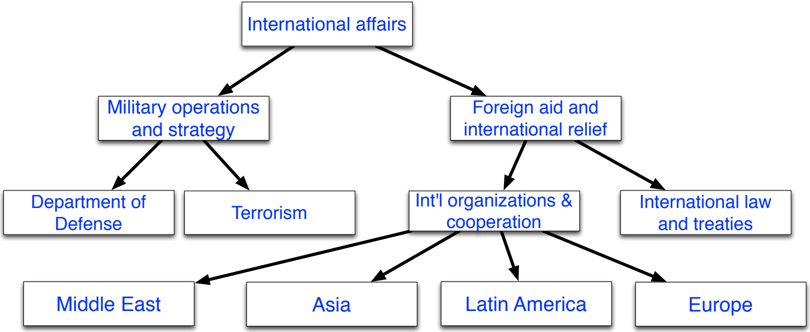


Approach Overview

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

Approach Overview

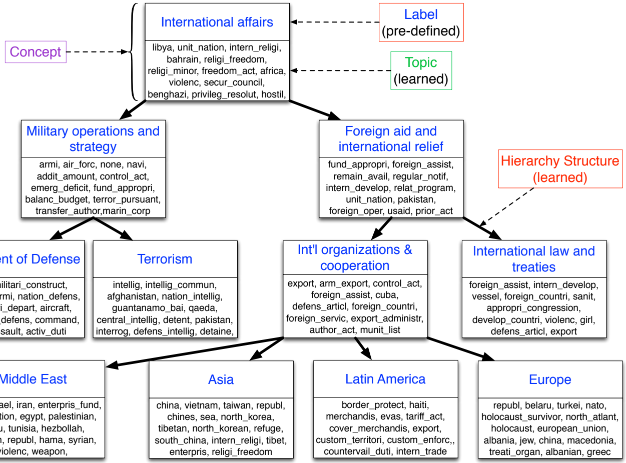
- Tree-structured hierarchy: captures label dependency and handles missing labels
- One topic per label: improves interpretability



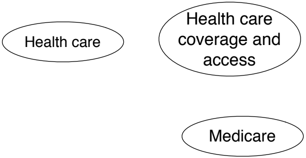
Approach Overview

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

L2H: Label-to-Hierarchy

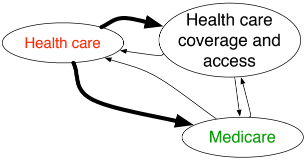


L2H Generative Process
1. Generating topic tree



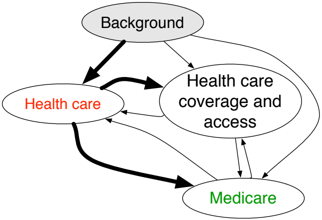
L2H Generative Process

1. Generating topic tree



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

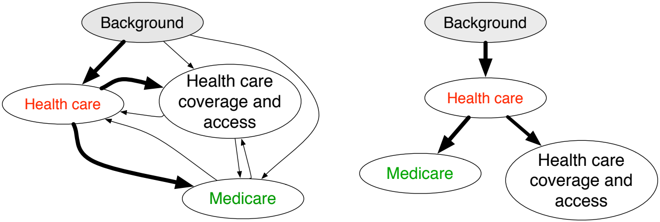
$$t_{\text{Health care}\rightarrow\text{Medicare}} = \frac{\text{No. docs tagged with both Health care \& Medicare}}{\text{No. docs tagged with Medicare}}$$



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

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

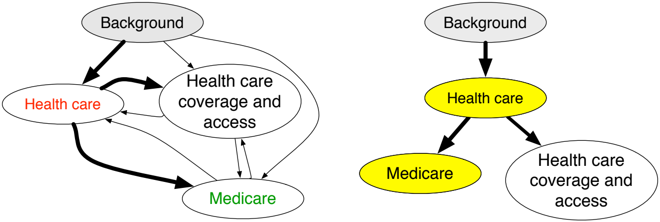
- Add a Background node to the graph



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



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

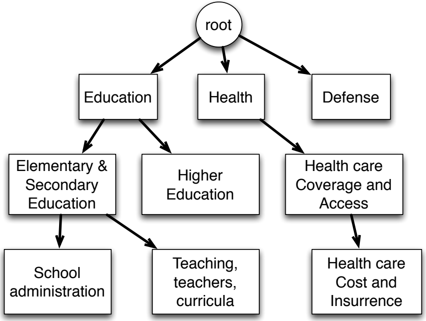
$$t_{\text{Health care}\to\text{Medicare}} = \frac{\text{No. docs tagged with both Health care \& Medicare}}{\text{No. docs tagged with Medicare}}$$

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

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

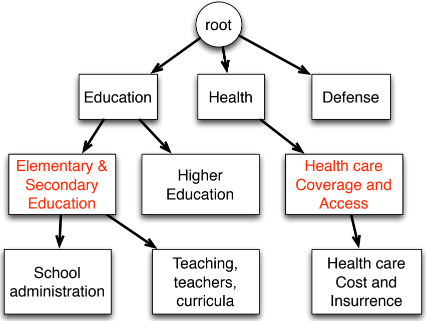
Given a document d with labels $\boldsymbol{l}_d$

- 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



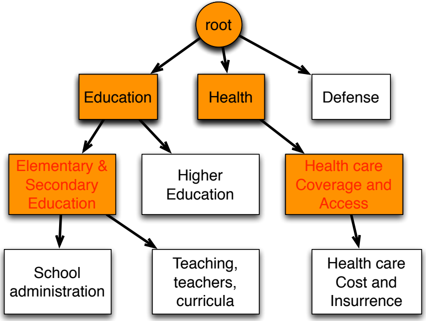
Given a document d with labels $\mathbf{l}_d$

- 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



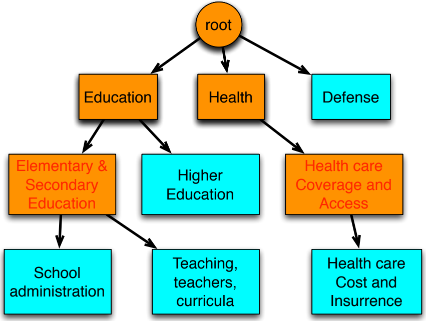
Given a document d with labels $\boldsymbol{l}_d$

- 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



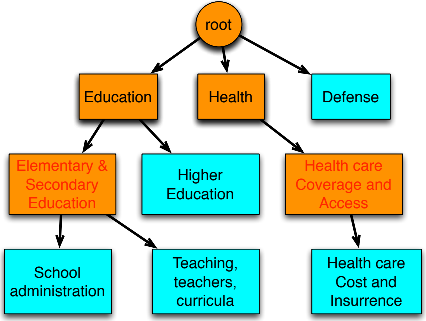
Given a document d with labels $\boldsymbol{l}_d$

- 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



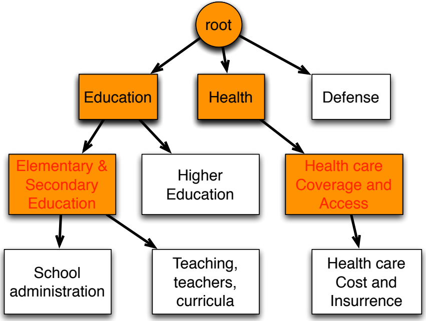
Given a document d with labels $\boldsymbol{l}_d$

- 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



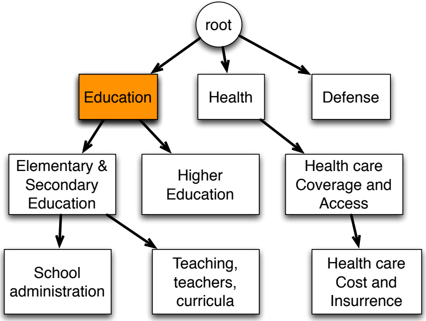
Given a document d with labels $\boldsymbol{l}_d$

- 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



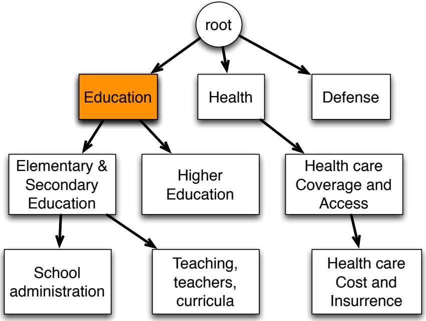
Given a document d with labels $\boldsymbol{l}_d$

- 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



Given a document d with labels $\boldsymbol{l}_d$

- 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

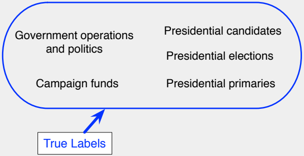


Multi-label Classification

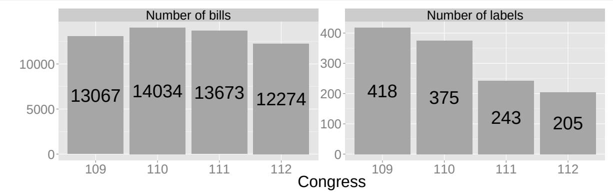
Task: Predicting multiple labels for each test document

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 (109^{th}-112^{th})

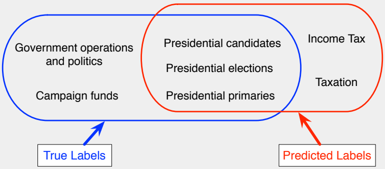


Multi-label Classification

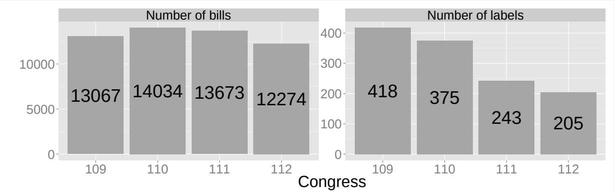
Task: Predicting multiple labels for each test document

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

Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features

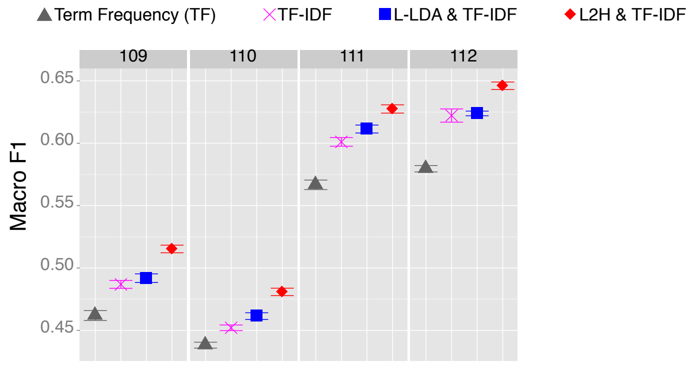
Multi-label Classification

Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features

▲ Term Frequency (TF) × TF-IDF ■ L-LDA & TF-IDF ◆ L2H & TF-IDF

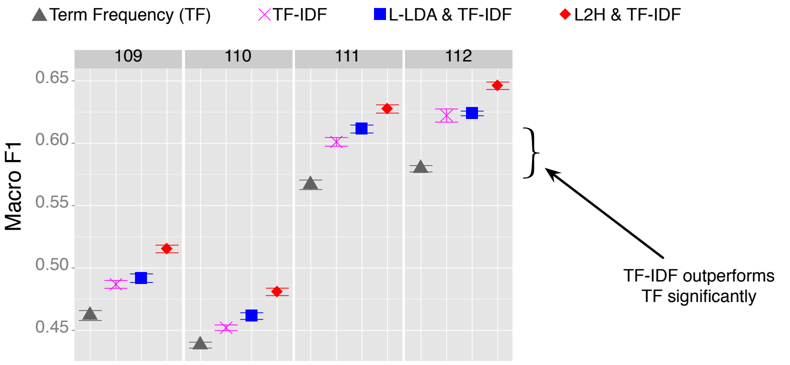
Multi-label Classification

Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features



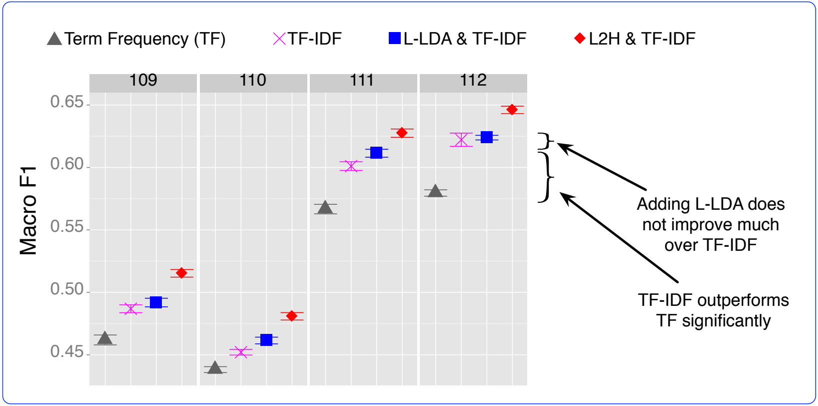
Multi-label Classification

Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features



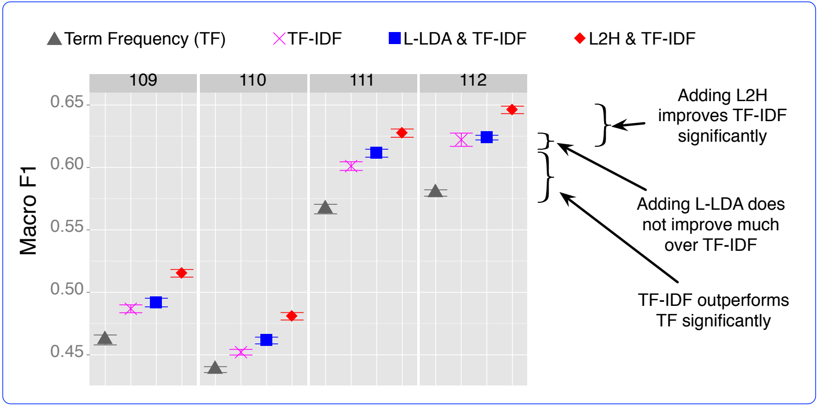
Multi-label Classification

Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features

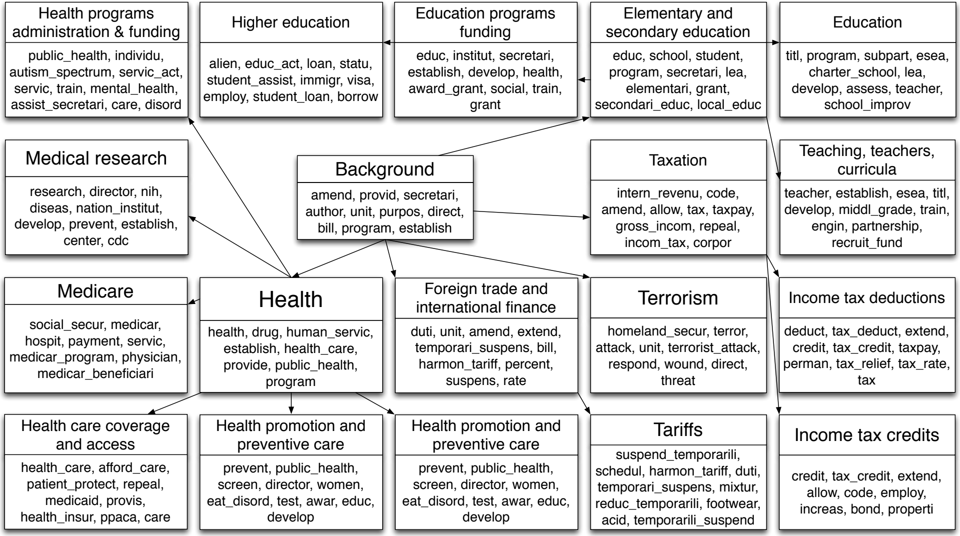


Multi-label Classification

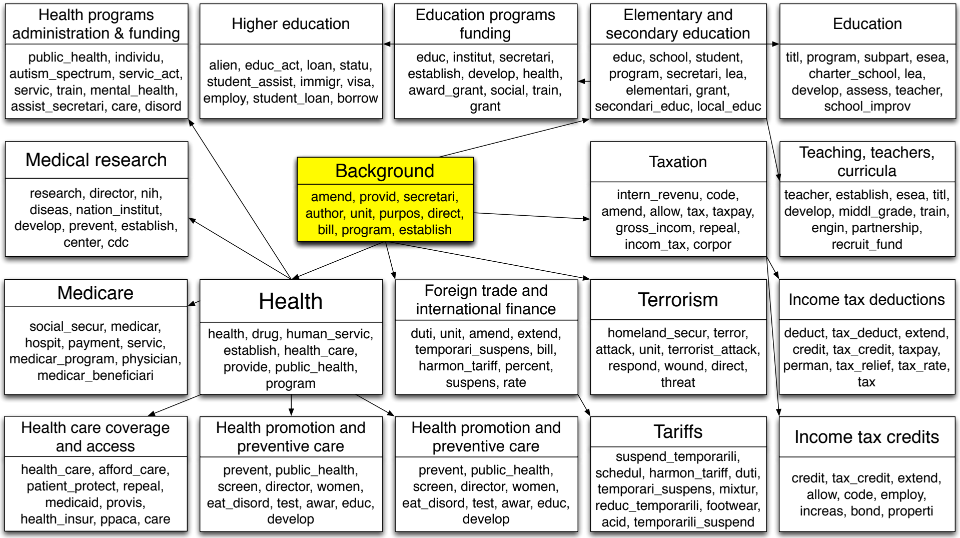
Method: Using M3L–an efficient max-margin multi-label classifier (Hariharan et al., MLJ'12) to study different sets of features



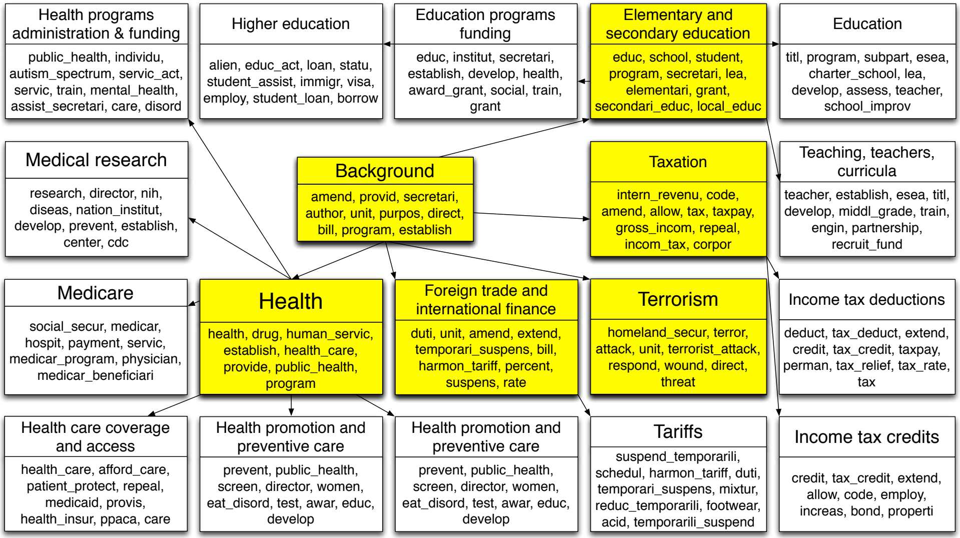
Partial Label Hierarchy learned from Congressional Bills



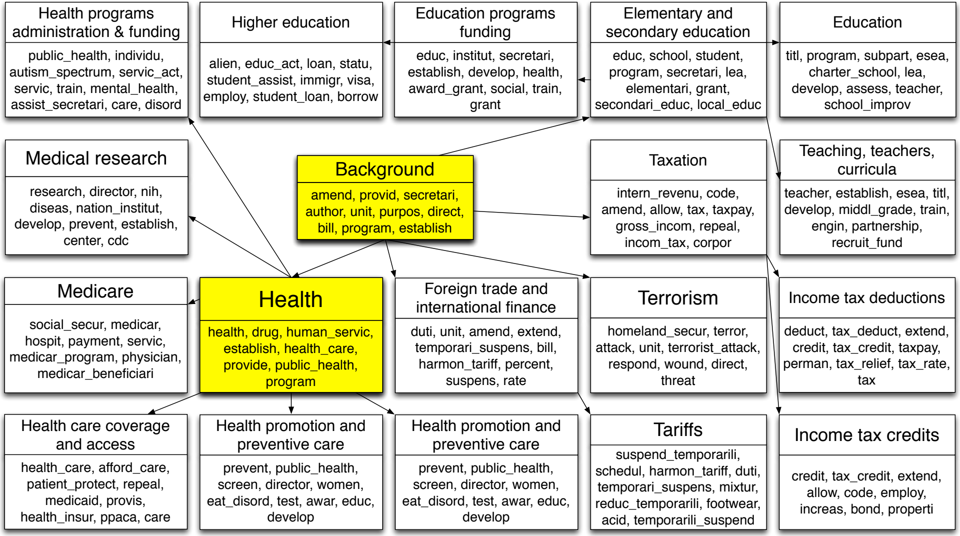
Partial Label Hierarchy learned from Congressional Bills



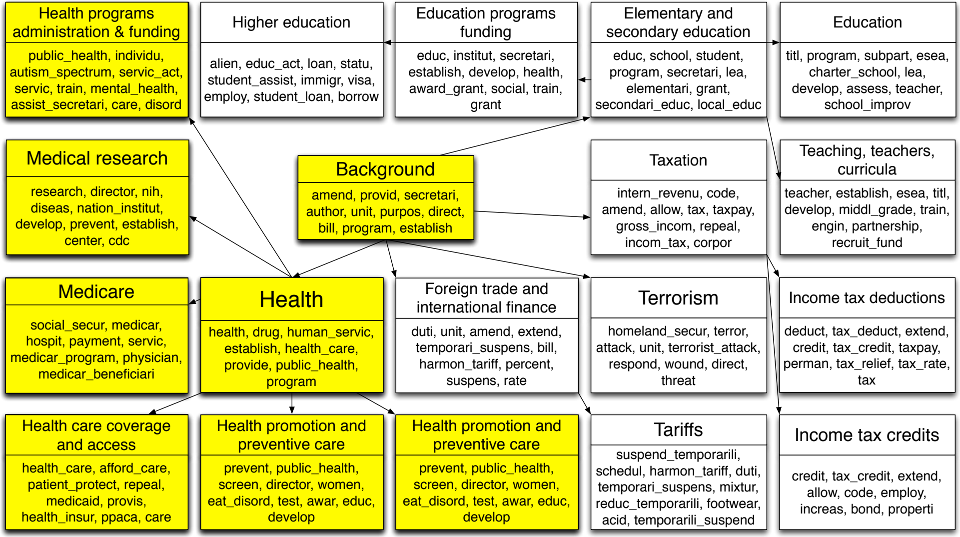
Partial Label Hierarchy learned from Congressional Bills



Partial Label Hierarchy learned from Congressional Bills



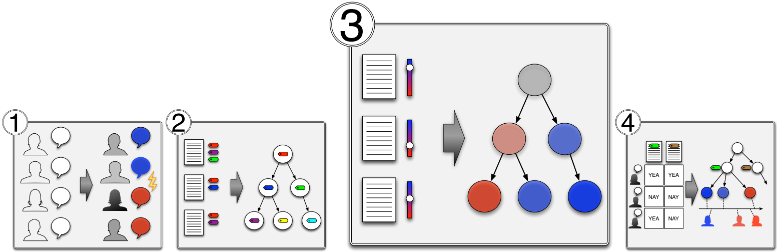
Partial Label Hierarchy learned from Congressional Bills



Summary

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



⋆ **V.-A. Nguyen**, J. Boyd-Graber, P. Resnik. *Lexical and Hierarchical Topic Regression* (**NIPS**, 2013)

Agenda-setting & Framing: Computational Approach

Agenda-setting

- **What** gets talked about?
- $\sim$ Topics

Agenda-setting & Framing: Computational Approach

Agenda-setting

- **What** gets talked about?
- $\sim$ Topics

Framing

- **How** things get talked about?
- $\sim$???

Agenda-setting & Framing: Computational Approach

Agenda-setting

- **What** gets talked about?
- $\sim$ Topics

Framing

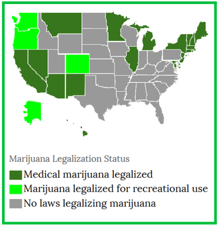
- **How** things get talked about?
- $\sim$???

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

Legalizing Marijuana

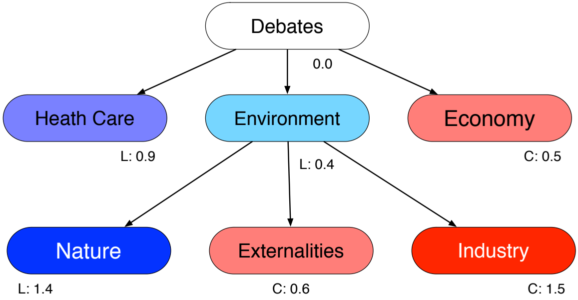
Economics

Health

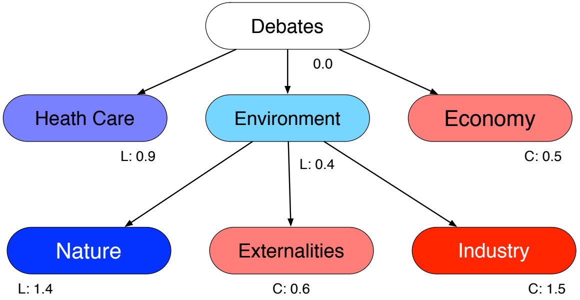


Legal process

Framing: Liberal vs. Conservative



Framing: Liberal vs. Conservative



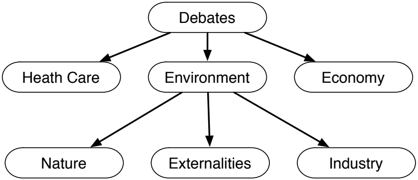
Input Data

- A collection of documents $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_D$
  - 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

Supervised Hierarchical Latent Dirichlet Allocation

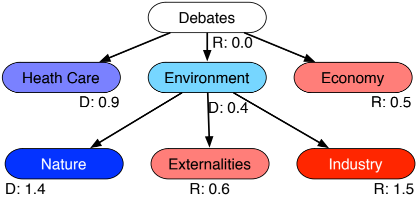
Modeling approach

- 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**



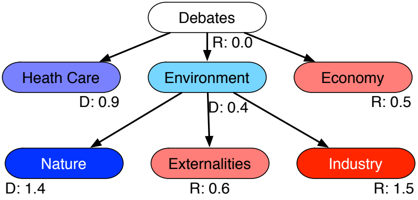
Modeling approach

- 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

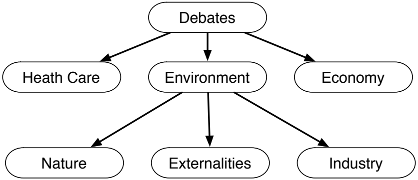


Modeling approach

- 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



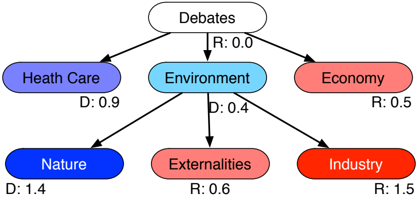
SHLDA: Generating words



For each node k in the tree

- Draw topic $\phi_k \sim \mathrm{Dir}(\beta)$

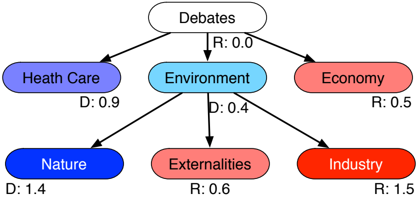
SHLDA: Generating words

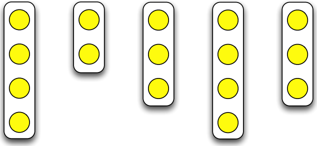


For each node k in the tree

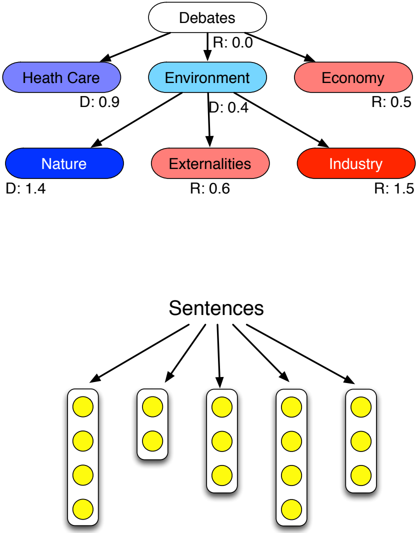
- Draw topic $\phi_k \sim \text{Dir}(\beta)$
- Draw regression parameter $\eta_k \sim \mathcal{N}(\mu, \sigma)$

SHLDA: Generating words

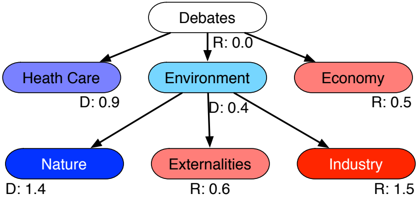


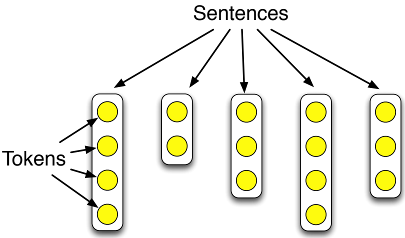


SHLDA: Generating words



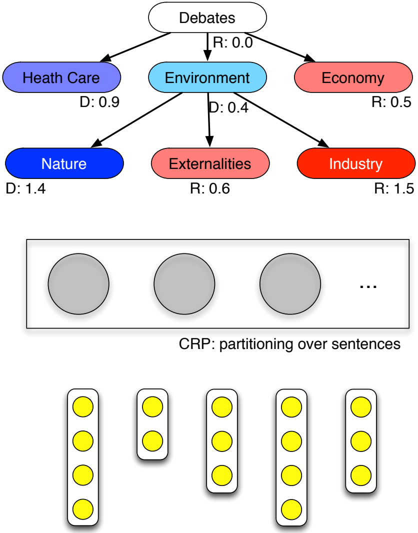
- Each document is a bag of sentences





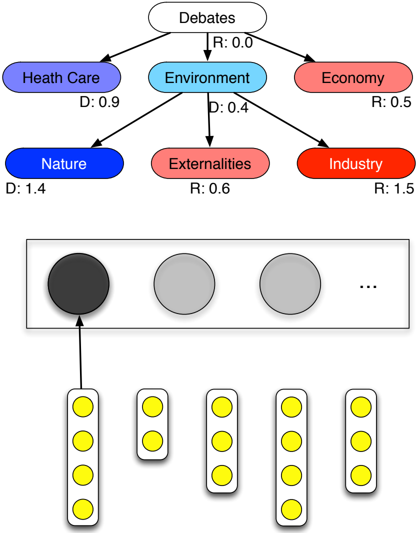
- Each document is a bag of sentences
- Each sentence is a bag of tokens

SHLDA: Generating words



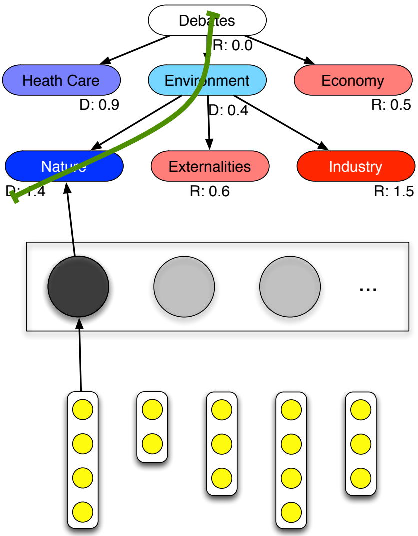
A Chinese restaurant process for each document to cluster similar sentences

SHLDA: Generating words

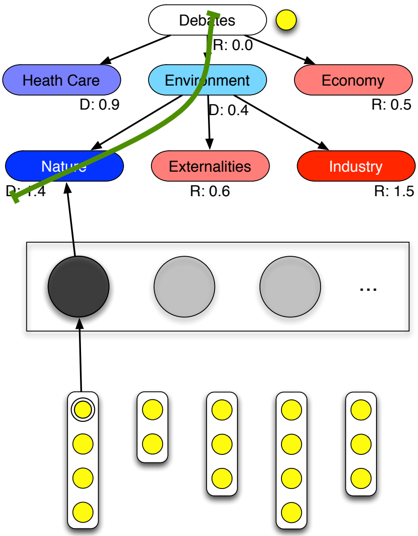


- Each sentence is assigned to a table

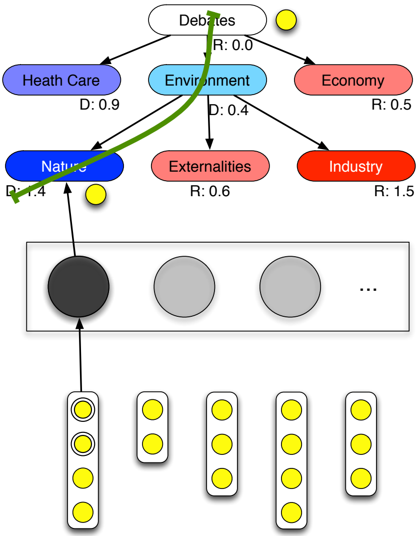
SHLDA: Generating words



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

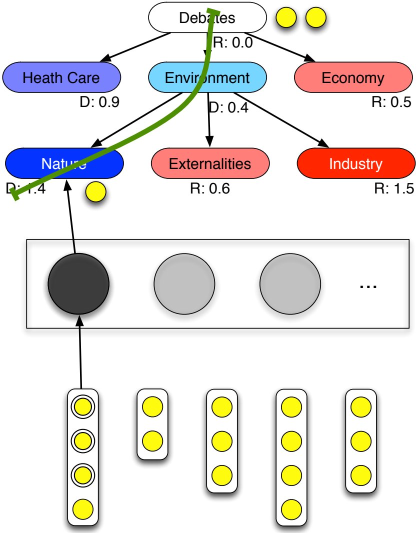


- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path



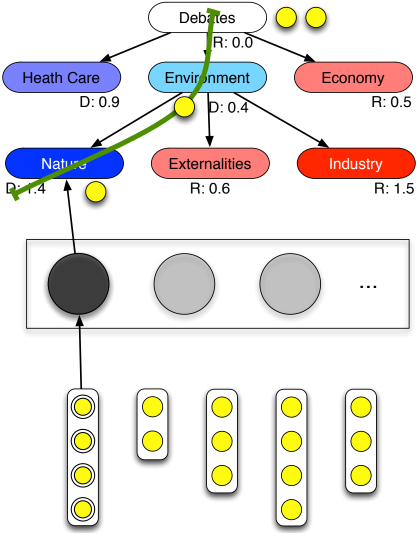
- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path

SHLDA: Generating words



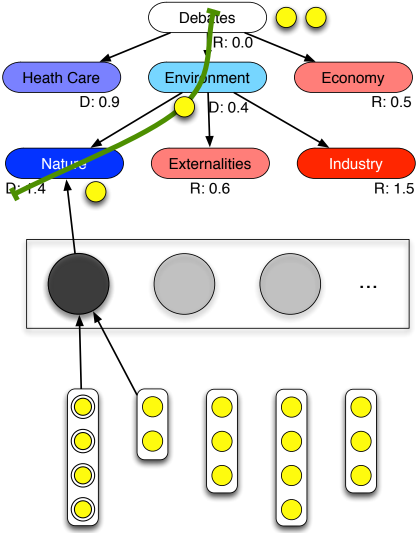
- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path

SHLDA: Generating words

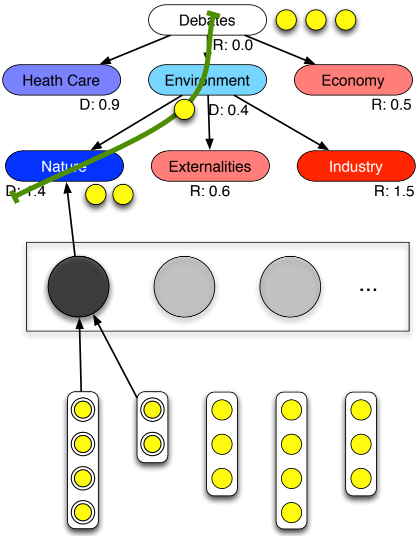


- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path

SHLDA: Generating words

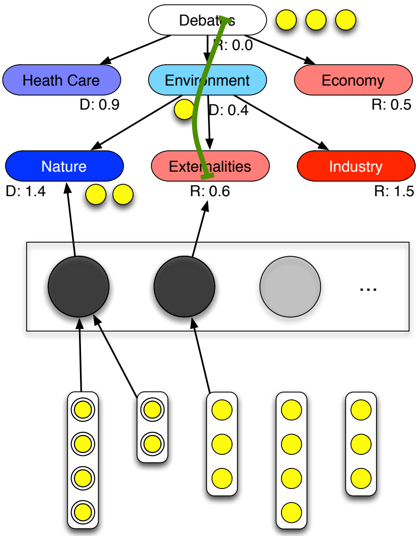


- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path



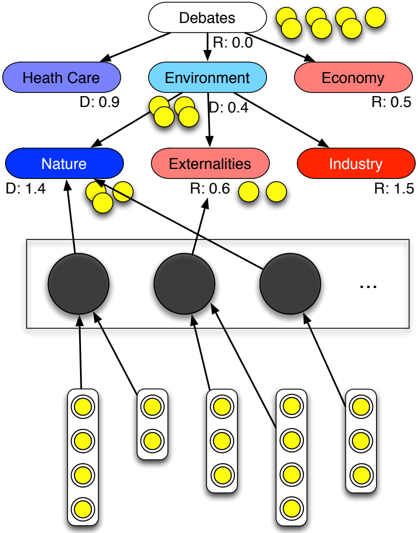
- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path

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



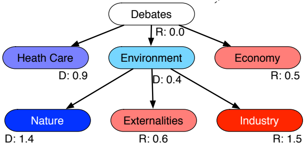
- Each sentence is assigned to a table
- Each table is assigned to a path
- Each token is assigned to a node on the chosen path

SHLDA: Generating response variable

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

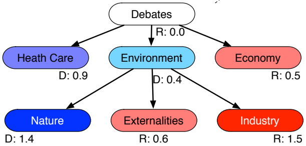
SHLDA: Generating response variable

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



SHLDA: Generating response variable

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

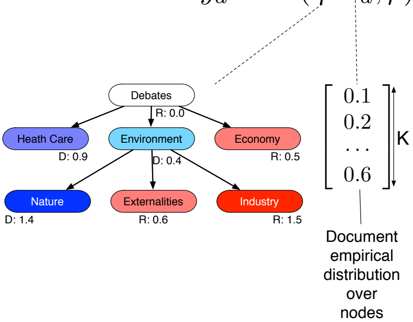


$$\begin{array}{cc} \begin{bmatrix} 0.1 \\ 0.2 \\ \cdots \\ 0.6 \end{bmatrix} & K \end{array}$$

Document empirical distribution over nodes

SHLDA: Generating response variable

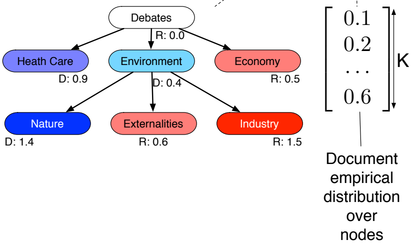
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho)$$



- Can capture issue-specific polarized words

SHLDA: Generating response variable

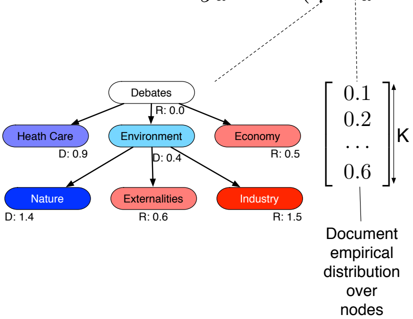
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho)$$



- 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

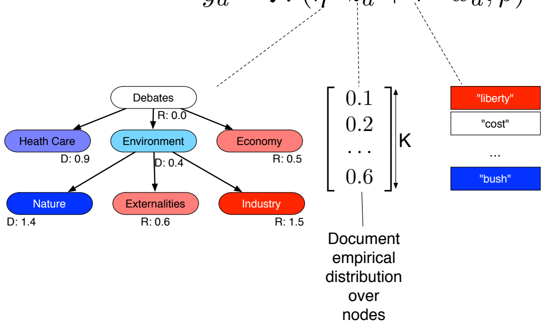
SHLDA: Generating response variable

$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

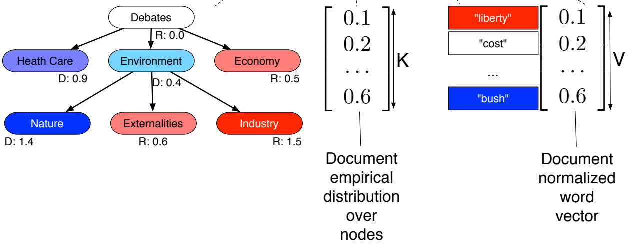
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

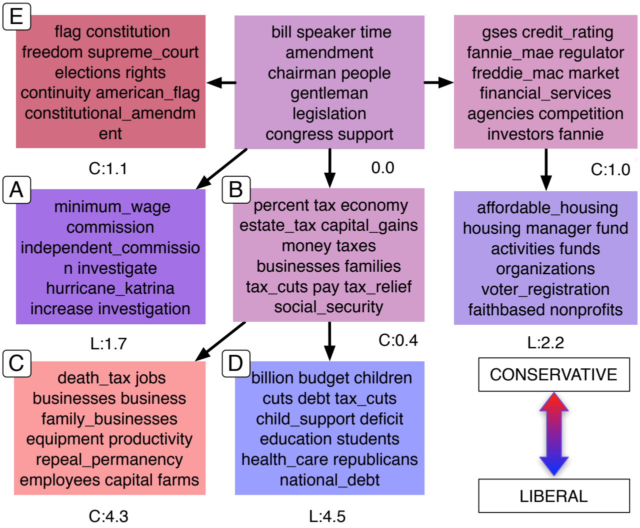
SHLDA: Generating response variable

$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

Qualitative results



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

- 109$^{\text{th}}$ congressional floor debates
- Amazon reviews
- Movie reviews

Evaluation

Mean square error averaged over 5 folds (lower is better)

Quantitative results: Regression

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

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

Quantitative results: Regression

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

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

Quantitative results: Regression

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

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

Quantitative results: Regression

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

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

Quantitative results: Regression

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

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

Quantitative results: Regression

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

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

Quantitative results: Regression

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

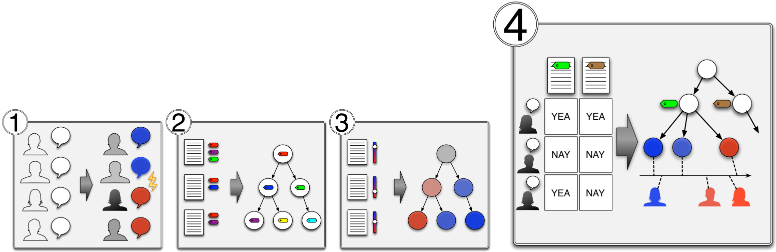
Table: Mean squared error averaged over 5 folds (lower is better).

Summary

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

Multi-dimensional Ideal Points from Votes and Text



- ⋆ 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

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

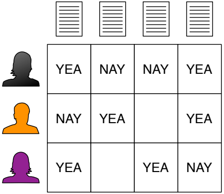
- People might have different positions on different issues
- $\rightarrow$ Multi-dimensional ideal points

One-dimensional Ideal Point using Votes

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

[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

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

One-dimensional ideal point
of legislator a

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

Polarity of bill b

Popularity of bill b

[Poole and Rosenthal, 1985]

Legislator a votes 'Yea' on bill b with probability

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

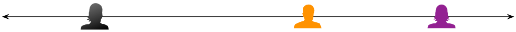
One-dimensional ideal point of legislator a

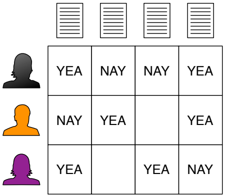
$$p(v_{a,b} = \text{Yea}) = \Phi(u_a x_b + y_b)$$

Polarity of bill b

Popularity of bill b

[Poole and Rosenthal, 1985]





YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

Legislator a votes 'Yea' on bill b with probability

One-dimensional ideal point of legislator a

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

Polarity of bill b

Popularity of bill b

[Poole and Rosenthal, 1985]

Legislator a votes 'Yea' on bill b with probability

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

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

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]

Legislator a votes 'Yea' on bill b with probability

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

Multi-dimensional ideal point
of legislator a

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

K dimensions

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]

Legislator a votes 'Yea' on bill b with probability

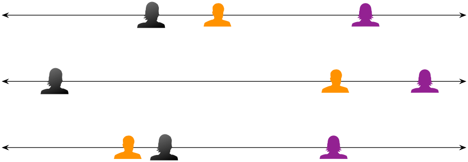
YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

Multi-dimensional ideal point of legislator a

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

K dimensions

[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]



Legislator a votes 'Yea' on bill b with probability

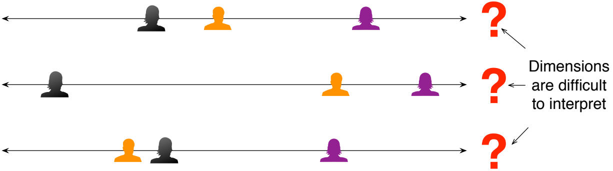
YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

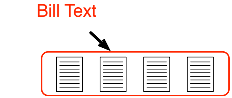
Multi-dimensional ideal point of legislator a

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

K dimensions

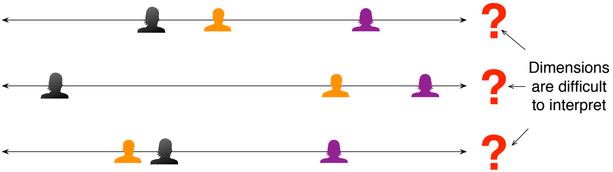
[Heckman and Jr., 1997, Jackman, 2001, Clinton et al., 2004]





YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

[Gerrish and Blei, 2012, Bonica, 2013, Lauderdale and Clark, 2014, Sim et al., 2015]



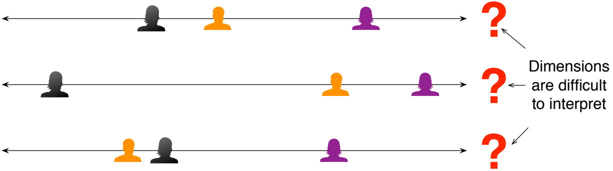
Multi-dimensional Ideal Point using Votes & **Text**

Legislator a votes 'Yea' on bill b with probability

YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

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

[Gerrish and Blei, 2012, Bonica, 2013, Lauderdale and Clark, 2014, Sim et al., 2015]



Multi-dimensional Ideal Point using Votes & **Text**

<u>Legislator a votes 'Yea' on bill b with probability</u>

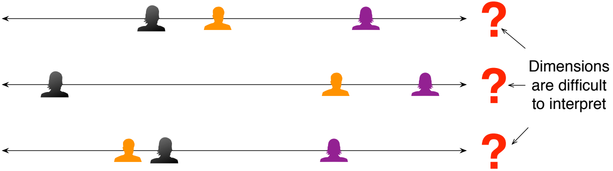
YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

Multi-dimensional ideal point of legislator a

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

Topic proportion of bill b estimated from its text

[Gerrish and Blei, 2012, Bonica, 2013, Lauderdale and Clark, 2014, Sim et al., 2015]



Legislator a votes 'Yea' on bill b with probability

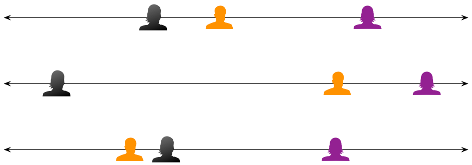
YEA	NAY	NAY	YEA
NAY	YEA		YEA
YEA		YEA	NAY

Multi-dimensional ideal point of legislator a

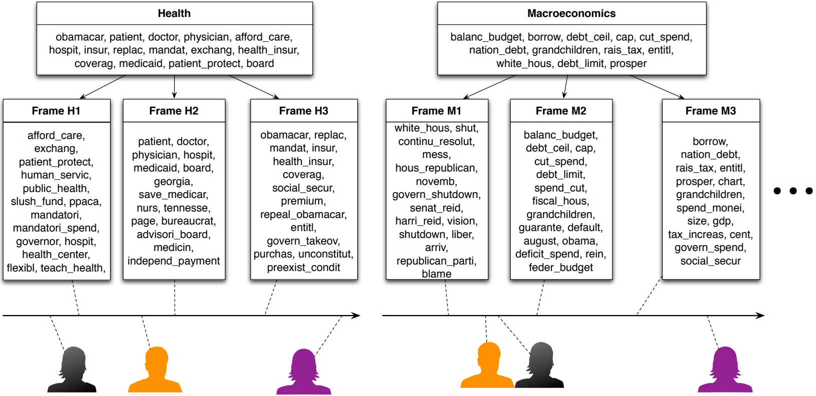
$$p(v_{a,b} = \text{Yea}) = \Phi\left(x_b \sum_{k=1}^{K} u_{a,k} \vartheta_{b,k} + y_b\right)$$

Topic proportion of bill b estimated from its text

[Gerrish and Blei, 2012, Bonica, 2013, Lauderdale and Clark, 2014, Sim et al., 2015]

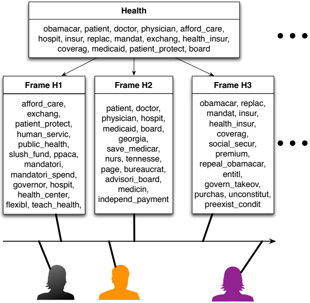


Our approach: Hierarchical Ideal Point Topic Model



Using both votes and text to learn

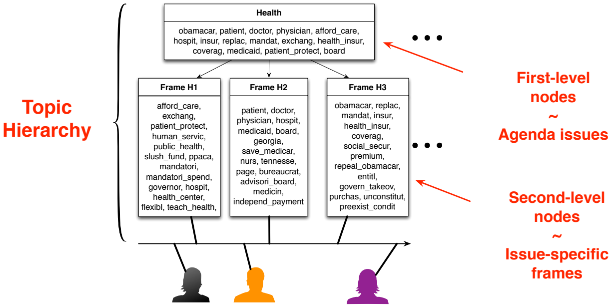
- Two-level topic hierarchy
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
  - Use existing labeled data to learn priors for interpretable issues
  - Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions



Approach Overview

Using both votes and text to learn

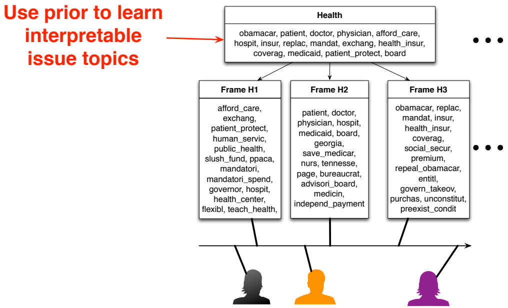
- Two-level topic hierarchy
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
  - Use existing labeled data to learn priors for interpretable issues
  - Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions



Approach Overview

Using both votes and text to learn

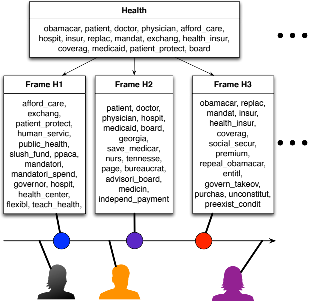
- Two-level topic hierarchy
  - First-level nodes map to agenda issues
  - Second-level nodes map to issue-specific frames
  - Use existing labeled data to learn priors for interpretable issues
  - Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions



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

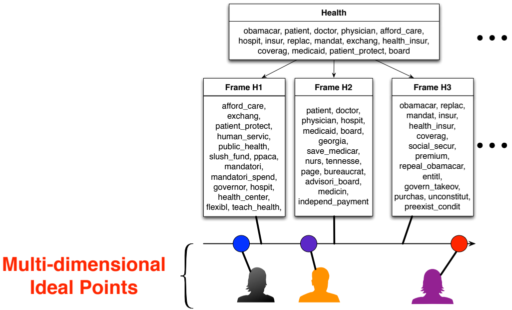
Learn ideal point for each frame



Approach Overview

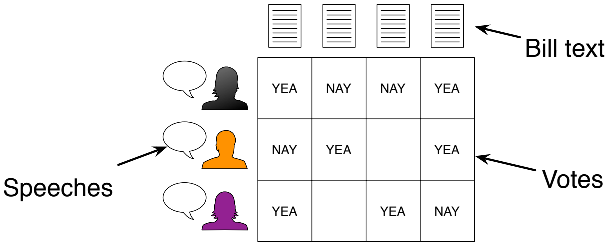
Using both votes and text to learn

- 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



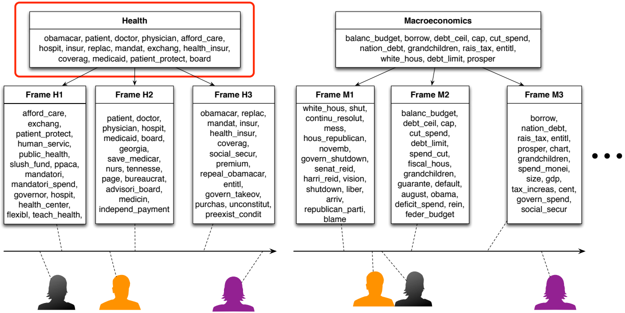
Inputs

- A collection of votes $\{v_{a,b}\}$
- A collection of D speeches $\{\boldsymbol{w}_d\}$, each of which is given by legislator a_d
- A collection of B bill text $\{\boldsymbol{w}'_b\}$



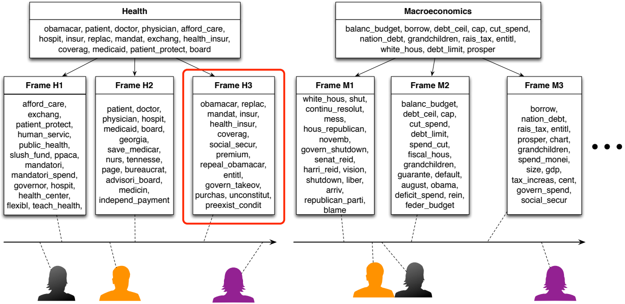
Modeling bill text

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



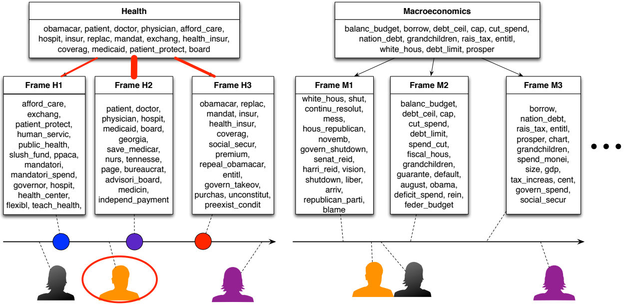
Modeling speeches

- Each speech d also has a distribution θ_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



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 $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$



Evaluation: Tea Party in the House

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]

Evaluation: Tea Party in the House

The Tea Party

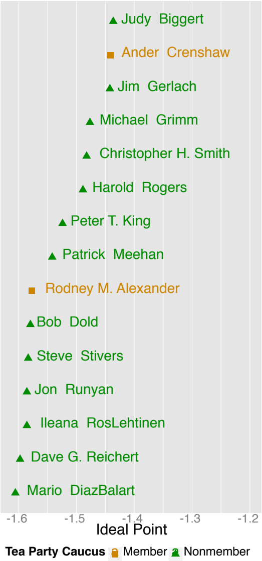
- Recent American political movement supporting more freedom, smaller government, lower tax
- Played an important role in recent electoral politics, especially within the Republican Party
- Organizations:
  - Institutional: Tea Party Caucus
  - Other: Tea Party Express, Tea Party Patriots, Freedom Works
- "**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 112$^{\text{th}}$ 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



One-dimensional Ideal Points

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

points and perspectives not offered elsewhere

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

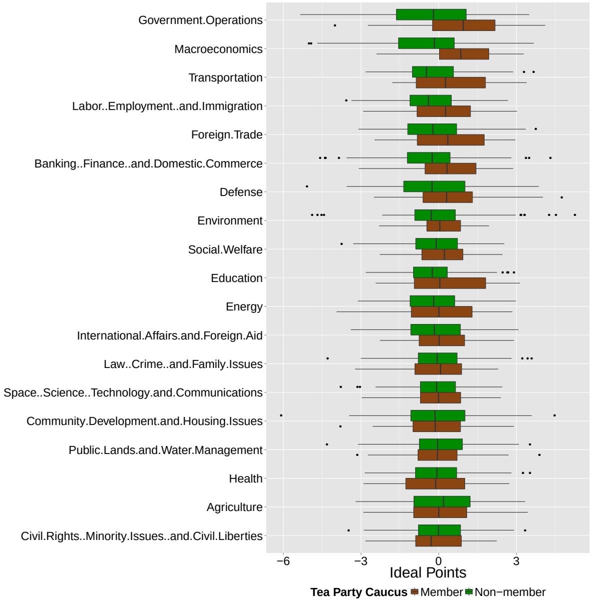
One-dimensional Ideal Points

- **Flake** and **Amash** didn't self-identify as members of the Tea Party Caucus but have been endorsed by other Tea Party organizations

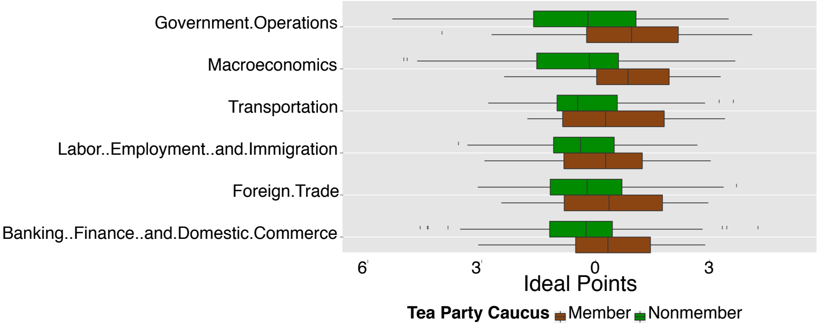
NEW REPUBLIC

"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."

Multi-dimensional Ideal Points



Multi-dimensional Ideal Points



Freedom Works' key votes on most highly polarized dimensions are about government spending

Tea Party Caucus Membership Prediction

Experiment setup

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

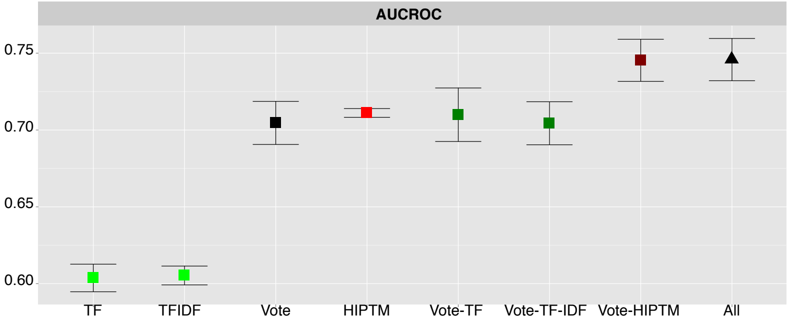
Tea Party Caucus Membership Prediction

Experiment setup

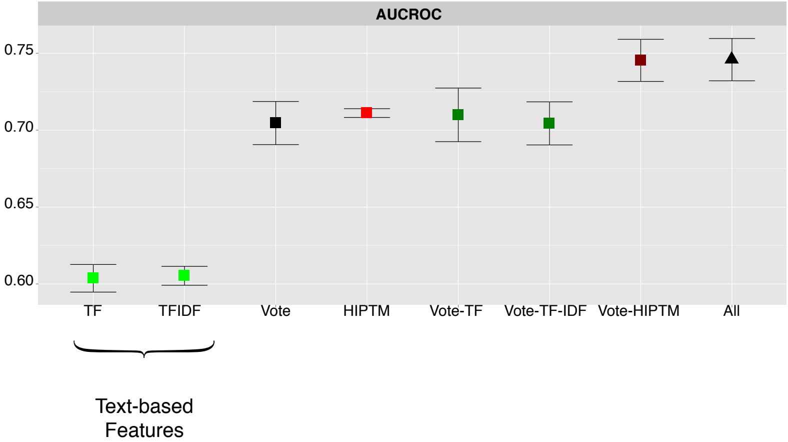
- Task: Binary classification of whether a legislator is a member of the Tea Party Caucus
- Evaluation metric: AUC-ROC
- Classifier: SVM^{light}
- Five-fold stratified cross-validation

Features

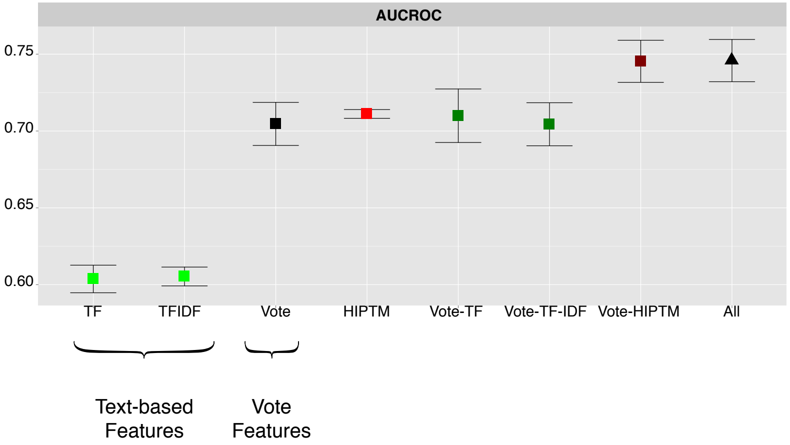
- Text-based features: normalized term frequency (**TF**) and **TF-IDF**
- **Vote**: binary features
- **HIPTM**: features extracted from our model including
  - K-dim ideal point $u_{a,k}$ estimated from both votes and text
  - K-dim ideal point estimated from text only $\boldsymbol{\eta}_k^T \hat{\psi}_{a,k}$
  - B probabilities estimating a's votes $\Phi(x_b \sum_{k=1}^{K} \vartheta_{b,k} u_{a,k} + y_b)$

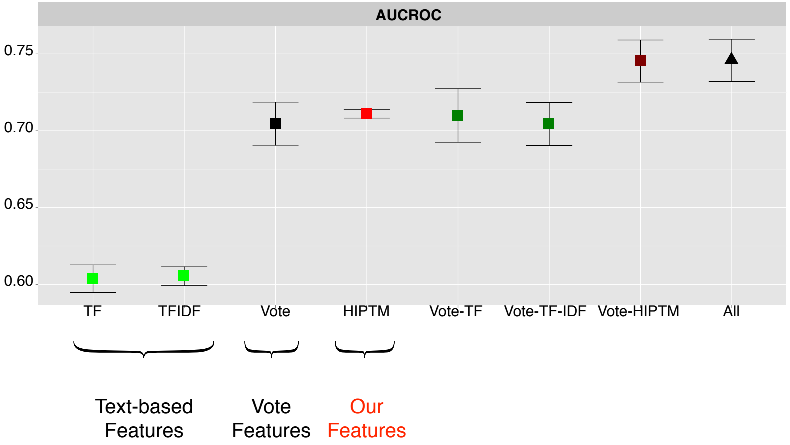

AUCROC
0.75
0.70
0.65
0.60
TF
TFIDF
Vote
HIPTM
Vote-TF
Vote-TF-IDF
Vote-HIPTM
All

Tea Party Caucus Membership Prediction: Votes & Text

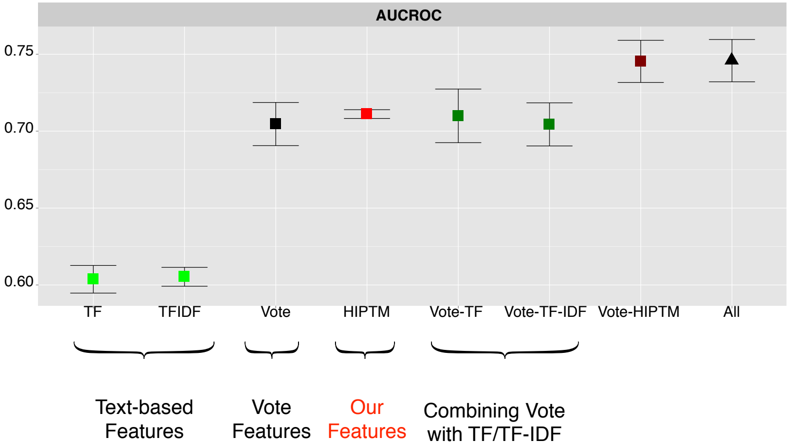


Tea Party Caucus Membership Prediction: Votes & Text

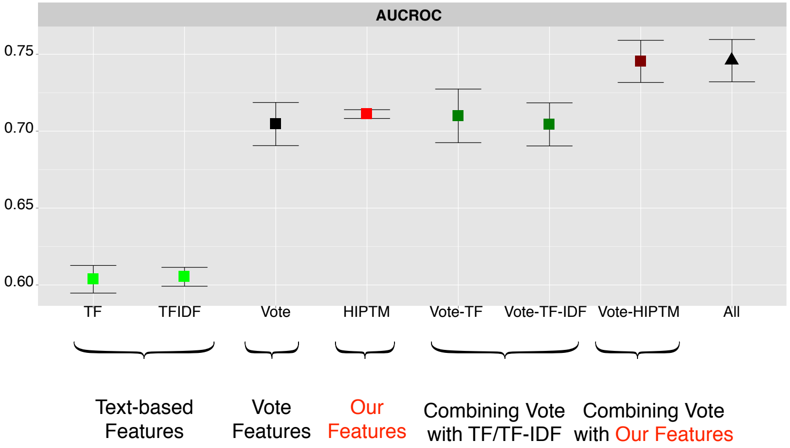



AUCROC
0.75
0.70
0.65
0.60
TF
TFIDF
Vote
HIPTM
Vote-TF
Vote-TF-IDF
Vote-HIPTM
All
Text-based Features
Vote Features
Our Features

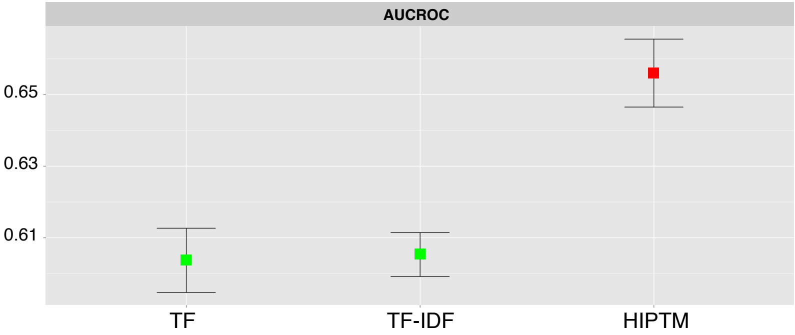
Tea Party Caucus Membership Prediction: Votes & Text



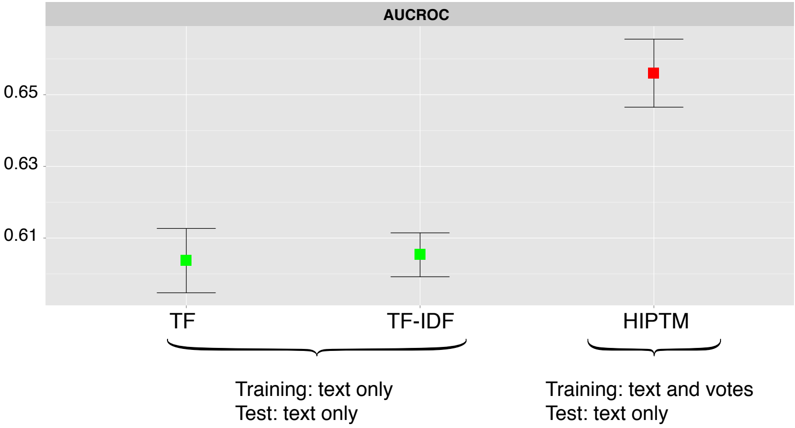
Tea Party Caucus Membership Prediction: Votes & Text



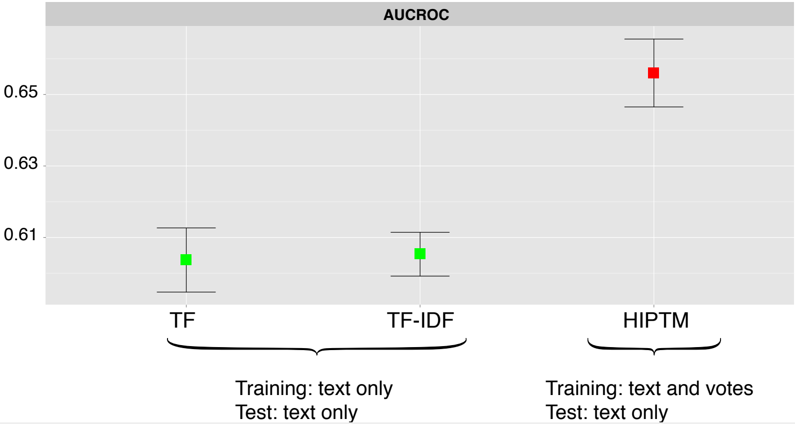
Tea Party Caucus Membership Prediction: Text Only



Tea Party Caucus Membership Prediction: Text Only



Tea Party Caucus Membership Prediction: Text Only



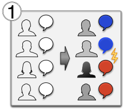
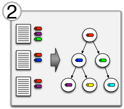
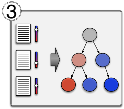
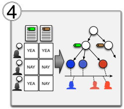
Vote-based features are not needed at test time, so this model makes it possible to do better prediction even for people who have no voting record in Congress

- e.g., new members of Congress or political candidates.

Summary

We introduce the *Hierarchical Ideal Point Topic Model* which extends existing multi-dimensional ideal point models using a hierarchy of topics, allowing us to

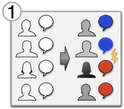
- Discover and analyze **agenda issues and issue-specific frames** in a unified framework
- Provide a formal computational model corresponding to the theory of **framing as second-level agenda setting**
- Analyze ideological positions of legislators in **multiple interpretable dimensions**
- Make **predictions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

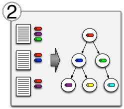





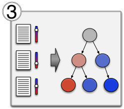
Technical Contributions

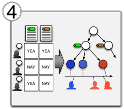
1. Extend prior work on topic segmentation in conversation by incorporating **speaker identity** and using **Bayesian nonparametrics**
2. Capture **dependency among labels** in multi-labeled data using a **tree-structured topic hierarchy**
3. Extend existing supervised topic model using a **hierarchy 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**

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**

Software & Data

Source code

- All introduced models: `https://github.com/vietansegan`

Data

- Congressional floor debates (109^{th}–112^{th}):
  `http://www.cs.umd.edu/~vietan/data/debates-109112.zip`
- *Crossfire* data:
  `http://www.cs.umd.edu/~vietan/topicshift/crossfire.zip`
- 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

SITS: Speaker Identity for Topic Segmentation

Problem

Study agendas and agenda control in political debates and other conversations

SITS

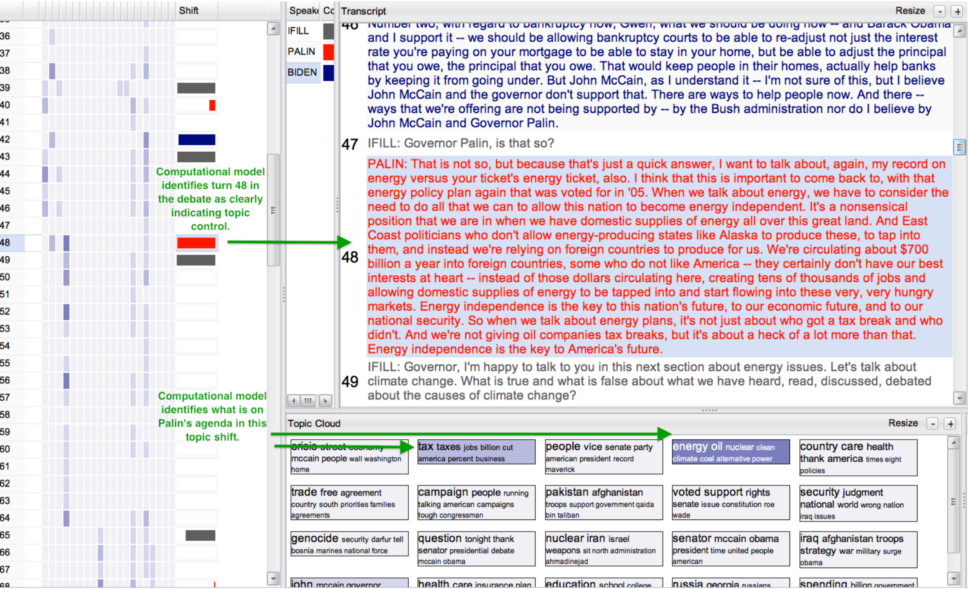
By modeling explicitly agenda control behaviors of speakers, SITS is able to discover

- 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

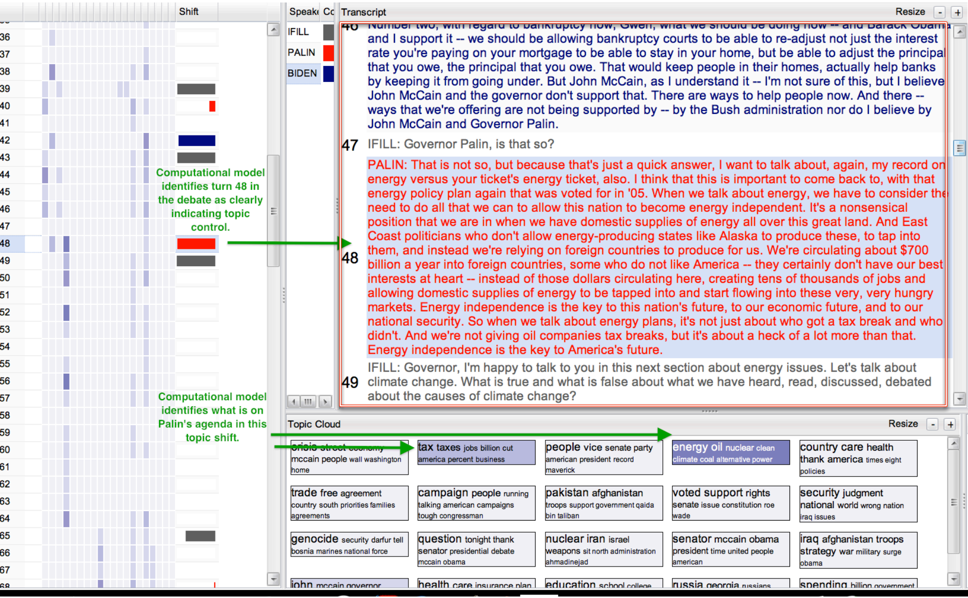
Applications

- 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

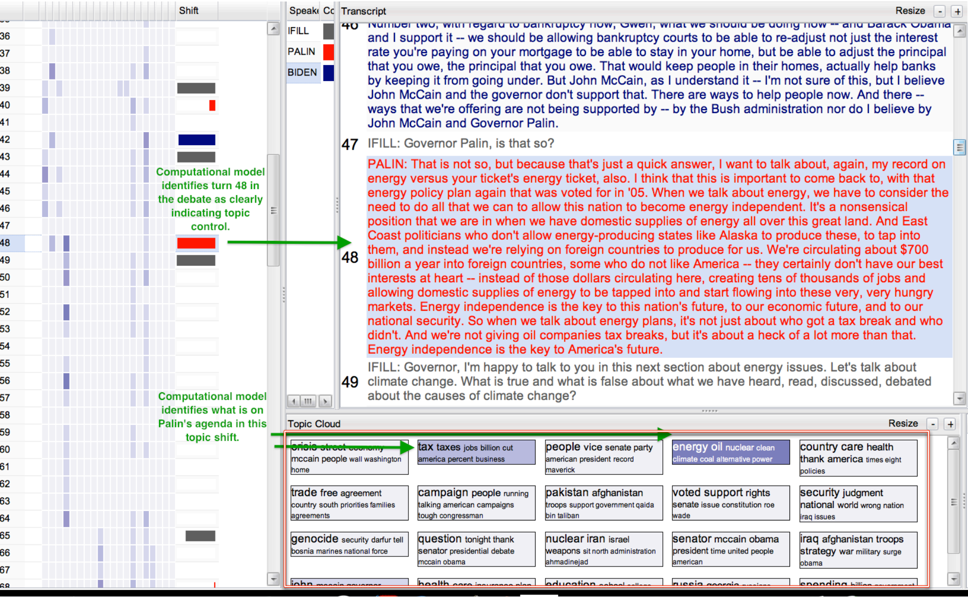
Argviz: Interactive Visualization of Topic Dynamics



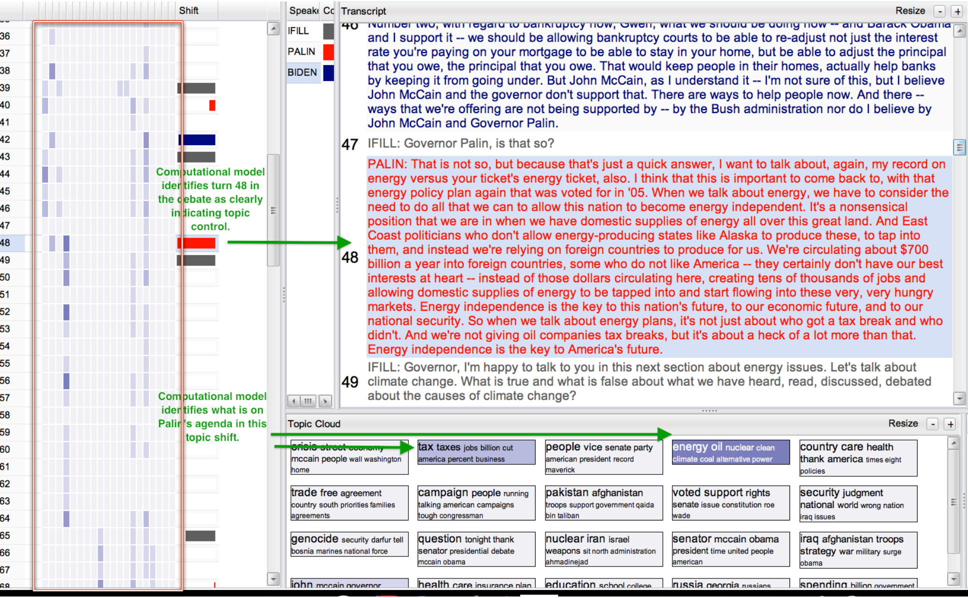
Argviz: Interactive Visualization of Topic Dynamics

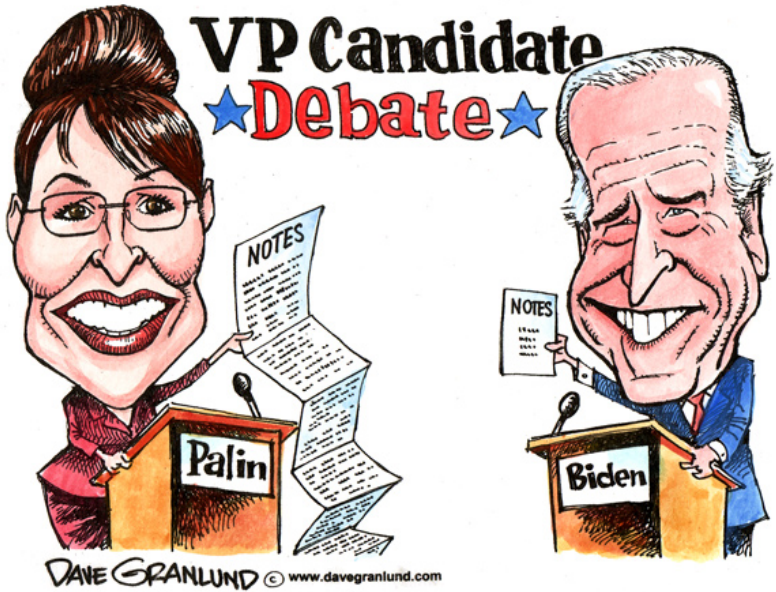


Argviz: Interactive Visualization of Topic Dynamics



Argviz: Interactive Visualization of Topic Dynamics




VP Candidate
Debate
NOTES
NOTES
Palin
Biden
DAVE GRANLUND © www.davegranlund.com

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.

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.

2008 U.S. Presidential Election

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.

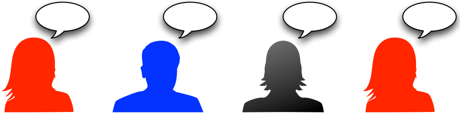
Joe Biden

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

Sarah Palin

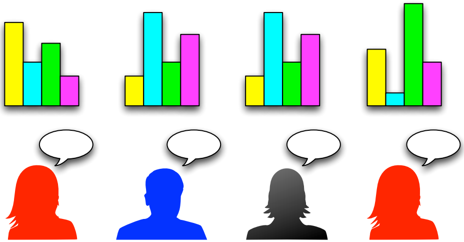
That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy . . . When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent . . . East Coast politicians who don't allow energy-producing 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



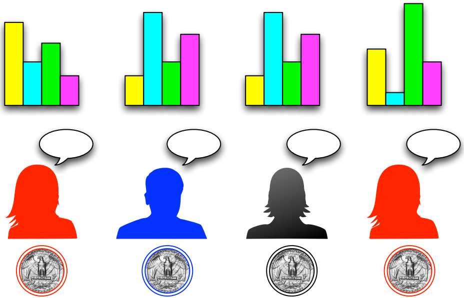
SITS: Speaker Identity for Topic Segmentation

Nonparametric Bayesian topic model for agenda control



- Each turn: a **multinomial distribution over topics**

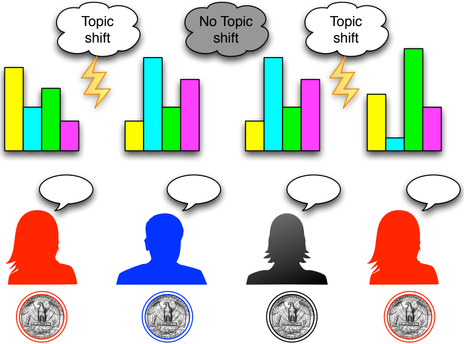
Nonparametric Bayesian topic model for agenda control



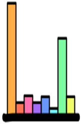
- Each turn: a **multinomial distribution over topics**
- Each speaker: a **biased coin** capturing how likely the speaker changes topic

SITS: Speaker Identity for Topic Segmentation

Nonparametric Bayesian topic model for agenda control



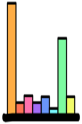
- Each turn: a **multinomial distribution over topics**
- Each speaker: a **biased coin** capturing how likely the speaker changes topic
- Each turn: a **binary latent variable** indicating whether the topic is shifted



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.

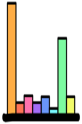
Generative Process



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.

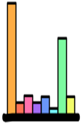
Generative Process



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.

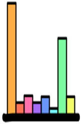
Generative Process



Gwen Ifill

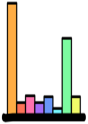
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

Generative Process

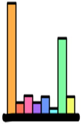


Gwen Ifill

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

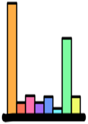


Generative Process



Gwen Ifill

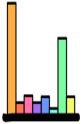
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.



Joe Biden

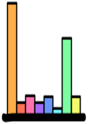
Well, mortgage-holders didn't pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .

Generative Process



Gwen Ifill

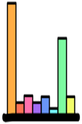
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.



Joe Biden

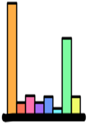
Well, mortgage-holders didn't pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .

Generative Process



Gwen Ifill

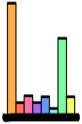
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.



Joe Biden

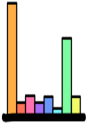
Well, mortgage-holders didn't pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .

Generative Process



Gwen Ifill

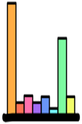
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.



Joe Biden

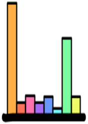
Well, mortgage-holders didn't pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .

Generative Process



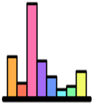
Gwen Ifill

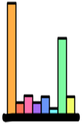
Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.



Joe Biden

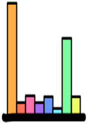
Well, mortgage-holders didn't pay the price. [...] Barack Obama pointed out two years ago that there was a subprime mortgage . . .





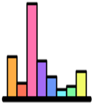
Gwen Ifill

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 . . .



Sarah Palin

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

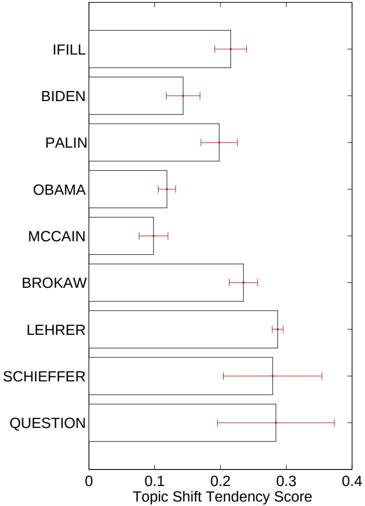
- the topic distribution of each conversational turn
- when the topic of the conversation changes
- how likely each speaker changes the topic of the conversation

Gibbs sampling

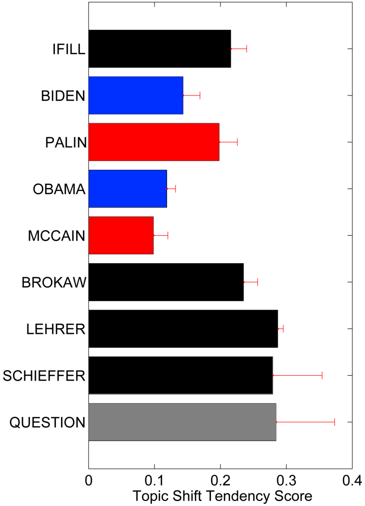
Alternates between

- **Sampling topic assignments**: which topic each token belongs to
- **Sampling topic shift indicator**: if topic shift occurs in each turn

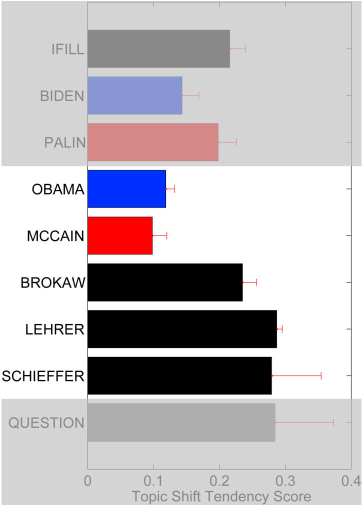
Agenda Control Behaviors in 2008 Presidential Debates



Agenda Control Behaviors in 2008 Presidential Debates

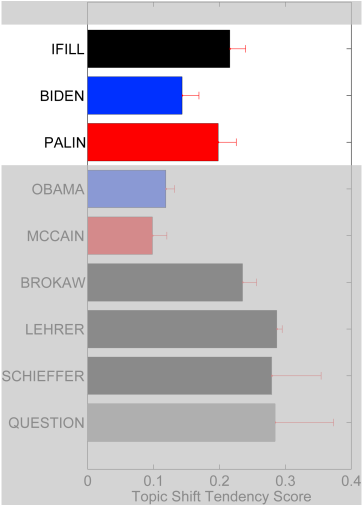


Agenda Control Behaviors in 2008 Presidential Debates

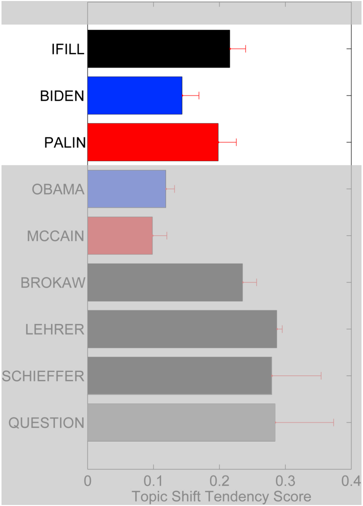


- In presidential debates, moderators have much higher scores than candidates do

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



- 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

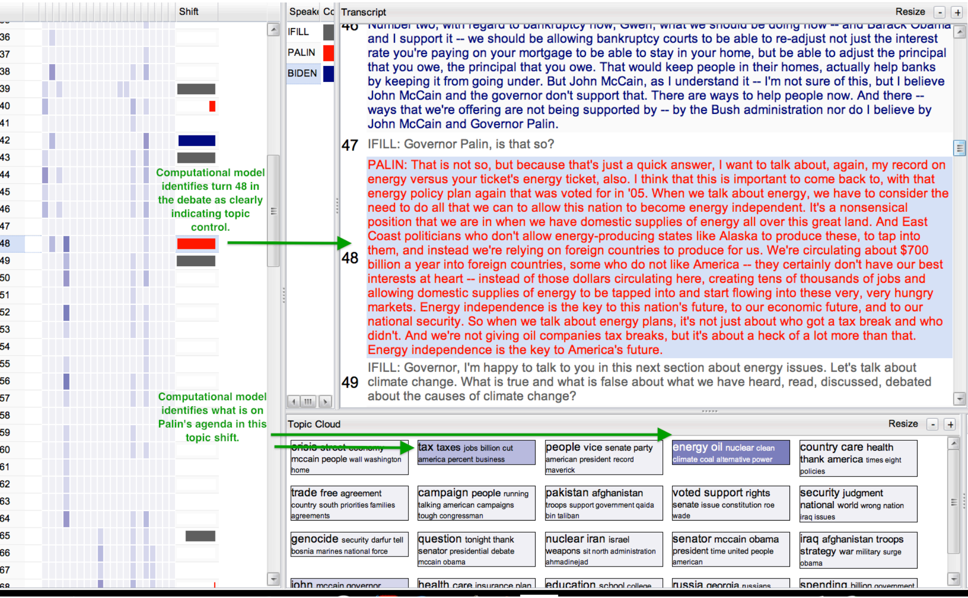
The Ifill Factor

By Scott Horton

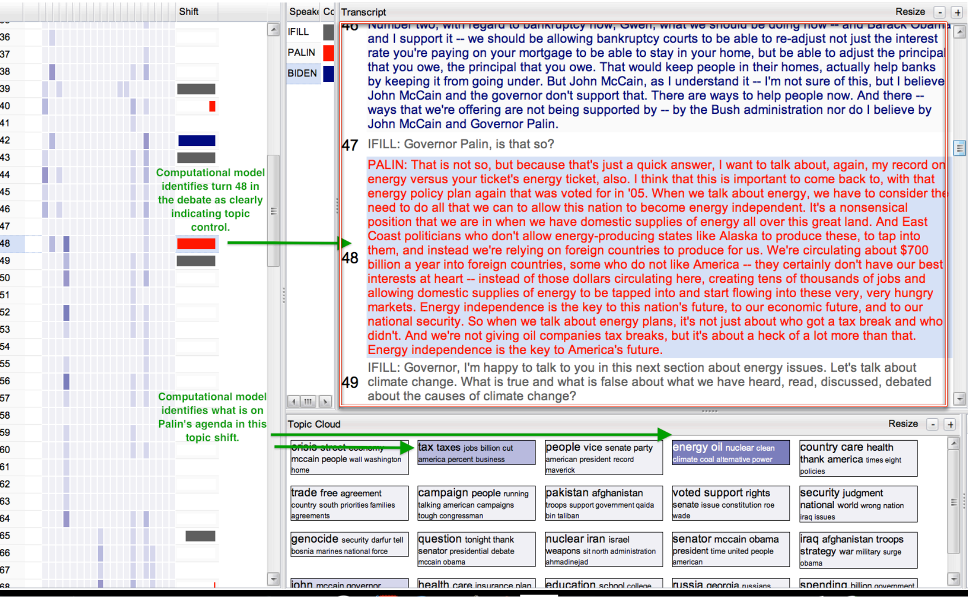
HARPER'S MAGAZINE

…. 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.

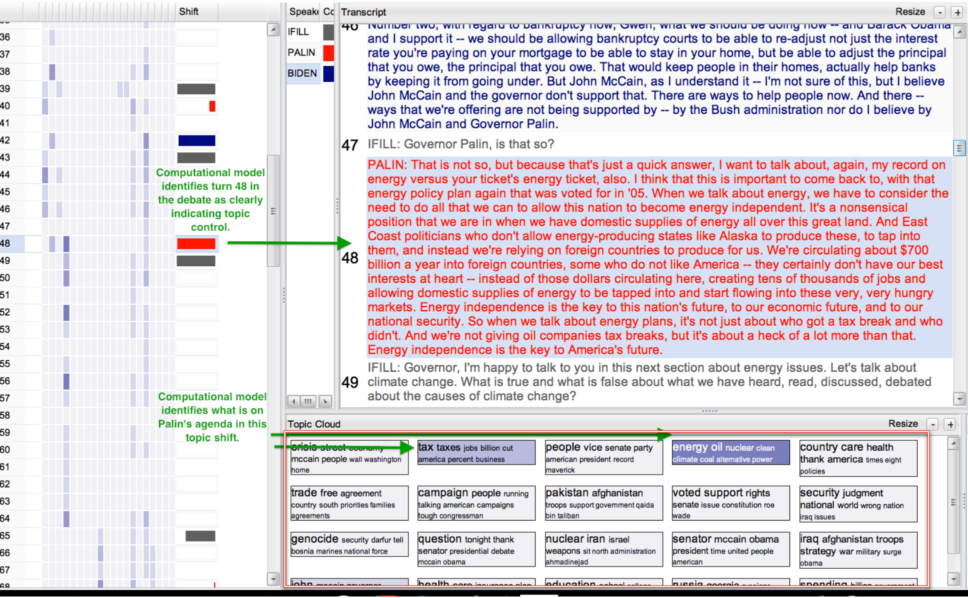
Argviz: Interactive Visualization of Topic Dynamics



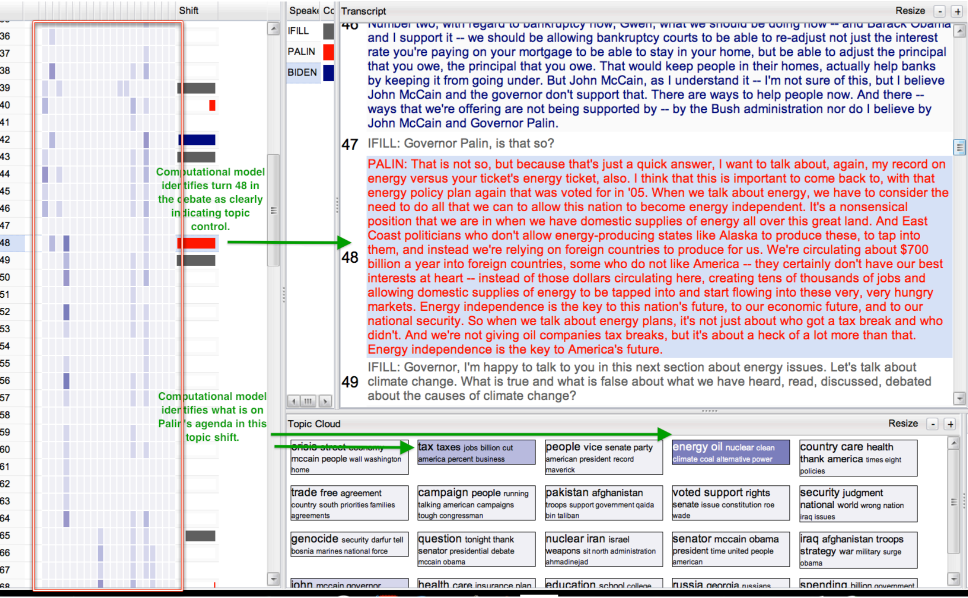
Argviz: Interactive Visualization of Topic Dynamics



Argviz: Interactive Visualization of Topic Dynamics



Argviz: Interactive Visualization of Topic Dynamics



Quantitative Evaluations

- Topic segmentation
- Influencer detection

Topic Segmentation

Task

Divide conversation into smaller, topically coherent segments

Topic Segmentation

Task

Divide conversation into smaller, topically coherent segments

Datasets

Datasets	Speakers	Conversations	Annotations	Content
ICSI Meetings	60	75	segmentation	engineering
2008 Debates	9	4	segmentation	politics

Topic Segmentation

Task

Divide conversation into smaller, topically coherent segments

Datasets

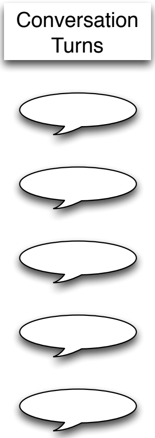
Datasets	Speakers	Conversations	Annotations	Content
ICSI Meetings	60	75	segmentation	engineering
2008 Debates	9	4	segmentation	politics

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

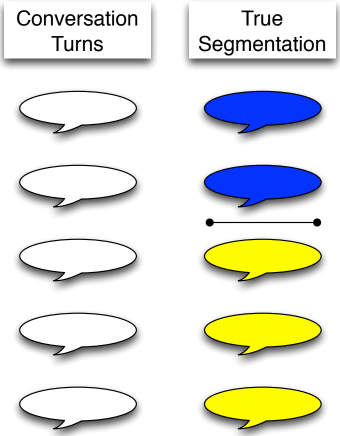
Topic Segmentation Evaluation: WindowDiff

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.



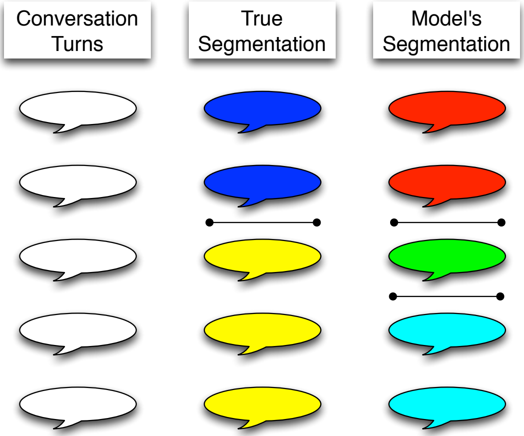
Topic Segmentation Evaluation: WindowDiff

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.



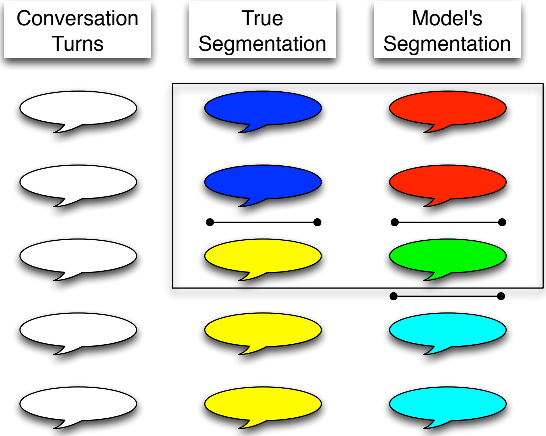
Topic Segmentation Evaluation: WindowDiff

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.



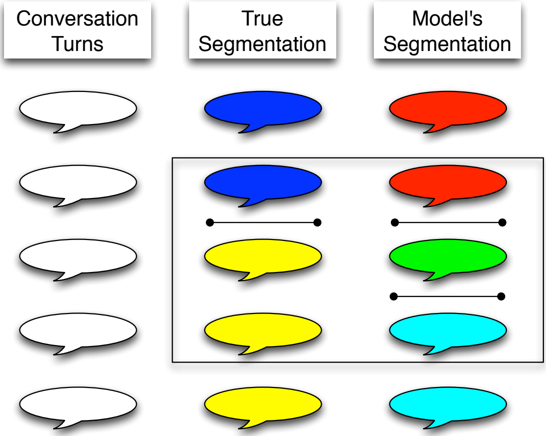
Topic Segmentation Evaluation: WindowDiff

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.



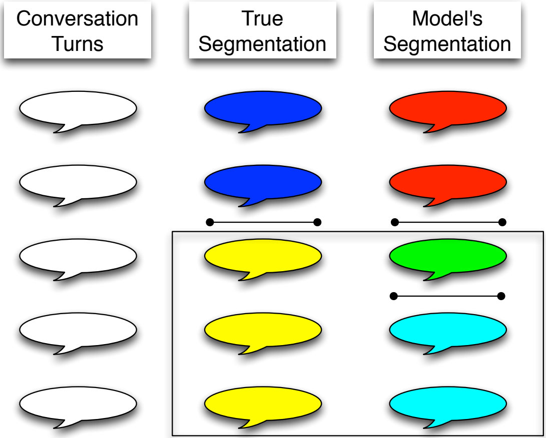
Topic Segmentation Evaluation: WindowDiff

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.

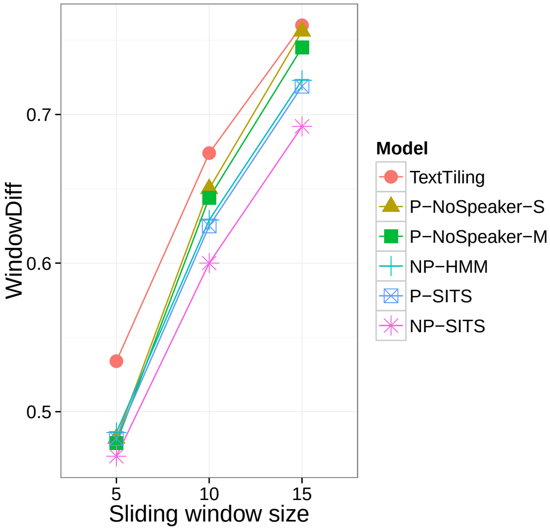


Topic Segmentation Evaluation: WindowDiff

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.



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

- 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

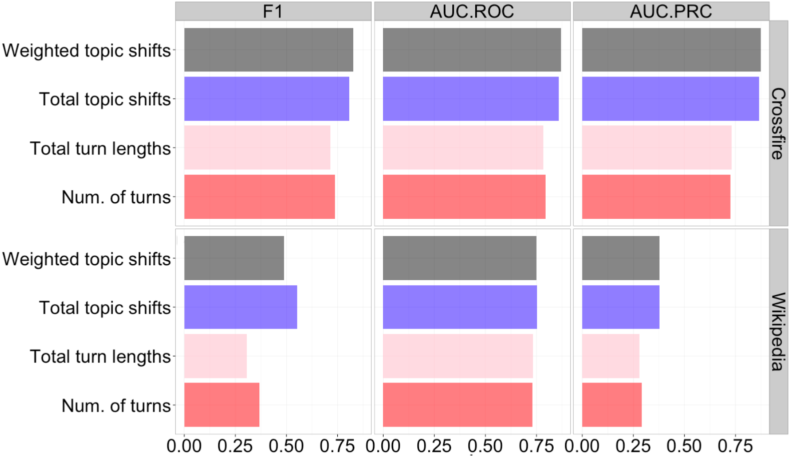
"the ability to change topical focus, especially given strong cultural and social pressure to be relevant, means having enough interpersonal power to take charge of the agenda"

[Palmer 1989]

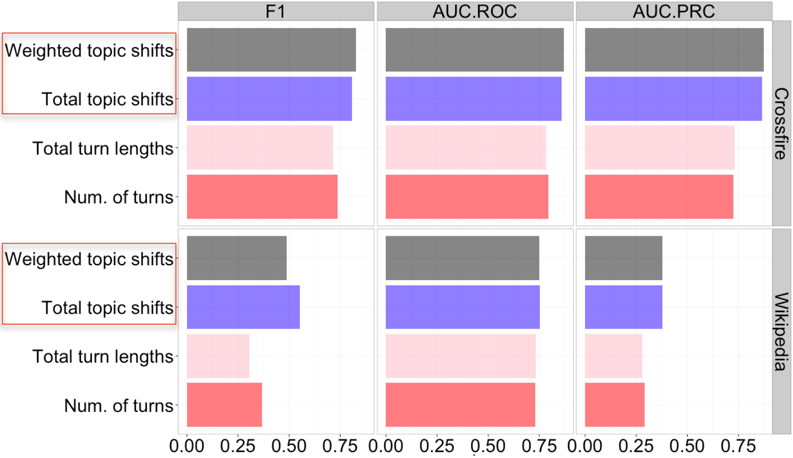
Datasets

Datasets	Speakers	Conversations	Annotations	Content
Crossfire	2567	1134	influencer	politics
Wikipedia	604	1991	influencer	varied

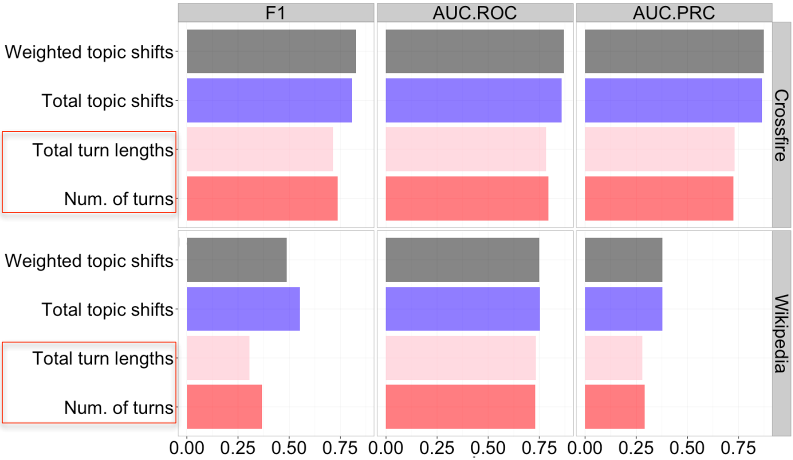
Influencer Detection



Influencer Detection



Influencer Detection



Summary

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”

Summary

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

Posterior Inference

Initialization

Initialize the label tree using the maximum spanning tree on $\mathcal{G}$ (Chu-Liu/Edmonds' algorithm)

MCMC Inference

Alternating between

1. Sampling the node assignment for each token
2. Sampling the topic ϕ 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

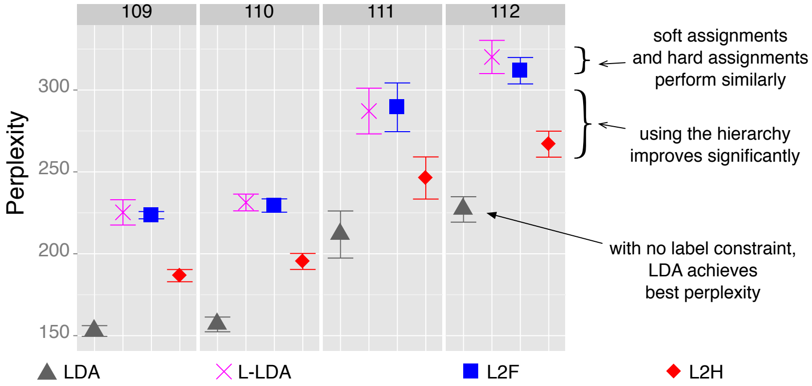
Initialization

Initialize the label tree using the maximum spanning tree on $\mathcal{G}$ (Chu-Liu/Edmonds' algorithm)

MCMC Inference

Alternating between

1. Sampling the node assignment for each token
2. Sampling the topic ϕ 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



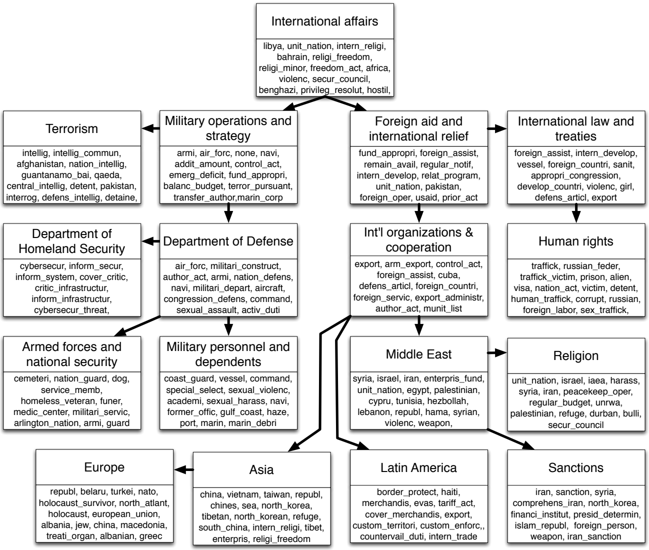
LDA
Unsupervised topic model, no label constraint

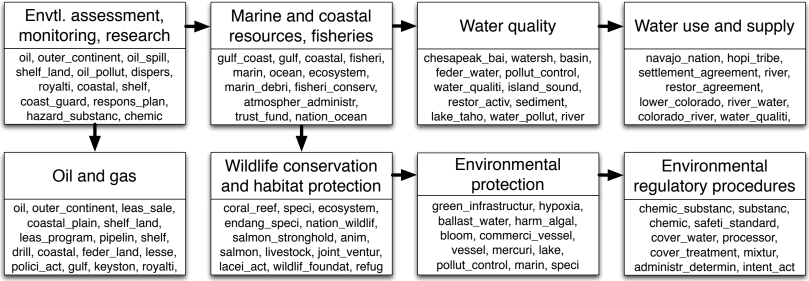
Labeled-LDA
(Ramage et al '09) one topic per label, flat structure, *hard assignment* (i.e., document can only be generated from its fixed set of topics)

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

Label-to-Hierarchy
Our proposed model

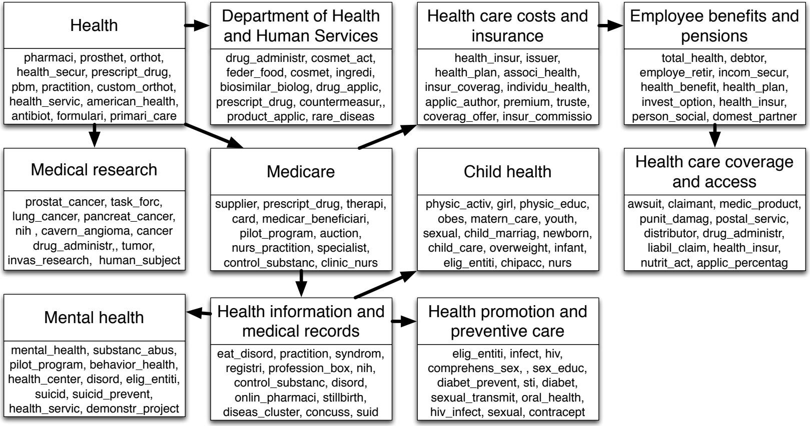
International Affairs





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
Water quality
chesapeak_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,
Oil and gas
oil, outer_continent, leas_sale, coastal_plain, shelf_land, leas_program, pipelin, shelf, drill, coastal, feder_land, lesse, polici_act, gulf, keyston, royalti,
Wildlife conservation and habitat protection
coral_reef, speci, ecosystem, endang_speci, nation_wildlif, salmon_stronghold, anim, salmon, livestock, joint_ventur, lacei_act, wildlif_foundat, refug
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

Health


Health
pharmaci, prosthet, orthot, health_secur, prescript_drug, pbm, practition, custom_orthot, health_servic, american_health, antibiot, formulari, primari_care
Department of Health and Human Services
drug_administr, cosmet_act, feder_food, cosmet, ingredi, biosimilar_biolog, drug_applic, prescript_drug, countermeasur,, product_applic, rare_diseas
Health care costs and insurance
health_insur, issuer, health_plan, associ_health, insur_coverag, individu_health, applic_author, premium, truste, coverag_offer, insur_commissio
Employee benefits and pensions
total_health, debtor, employe_retir, incom_secur, health_benefit, health_plan, invest_option, health_insur, person_social, domest_partner
Medical research
prostat_cancer, task_forc, lung_cancer, pancreat_cancer, nih , cavern_angioma, cancer drug_administr,, tumor, invas_research, human_subject
Medicare
supplier, prescript_drug, therapi, card, medicar_beneficiari, pilot_program, auction, nurs_practition, specialist, control_substanc, clinic_nurs
Child health
physic_activ, girl, physic_educ, obes, matern_care, youth, sexual, child_marriag, newborn, child_care, overweight, infant, elig_entiti, chipacc, nurs
Health care coverage and access
awsuit, claimant, medic_product, punit_damag, postal_servic, distributor, drug_administr, liabil_claim, health_insur, nutrit_act, applic_percentag
Mental health
mental_health, substanc_abus, pilot_program, behavior_health, health_center, disord, elig_entiti, suicid, suicid_prevent, health_servic, demonstr_project
Health information and medical records
eat_disord, practition, syndrom, registri, profession_box, nih, control_substanc, disord, onlin_pharmaci, stillbirth, diseas_cluster, concuss, suid
Health promotion and preventive care
elig_entiti, infect, hiv, comprehens_sex, , sex_educ, diabet_prevent, sti, diabet, sexual_transmit, oral_health, hiv_infect, sexual, contracept

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 $\rightarrow$ 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

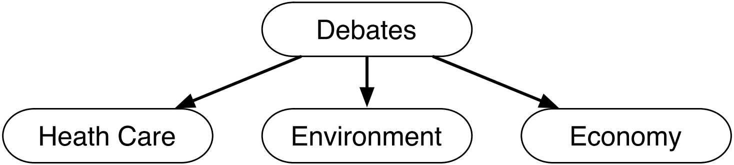
Framing: second-level agenda setting

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

Agenda setting & Framing

Framing: second-level agenda setting

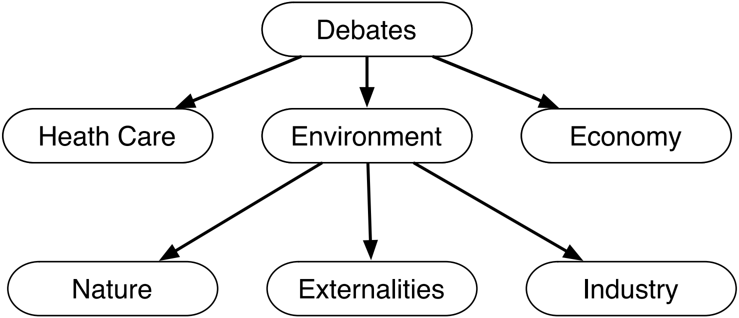
- Agenda setting: which **issues** are salient
- Framing: which **aspects** of the discussed issues are salient



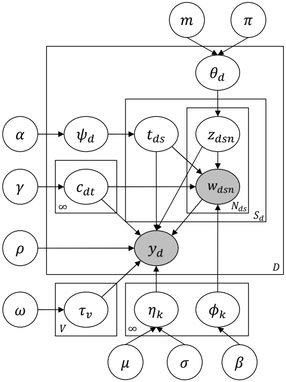
Agenda setting & Framing

Framing: second-level agenda setting

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



Supervised Hierarchical LDA (SHLDA)



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 $y_d \sim \mathcal{N}(\boldsymbol{\eta}^T \bar{\boldsymbol{z}}_d + \boldsymbol{\tau}^T \bar{\boldsymbol{w}}_d, \rho)$:
      i. $\bar{z}_{d,k} = \frac{1}{N_{d,\cdot}} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} \mathbb{I}\left[k_{d,s,n} = k\right]$
      ii. $\bar{w}_{d,v} = \frac{1}{N_{d,\cdot}} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} \mathbb{I}\left[w_{d,s,n} = v\right]$

Data

- A collection of documents $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_D$
- Each document d has an associated **response variable** y_d

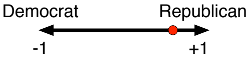
- A collection of documents $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_D$
- Each document d has an associated **response variable** y_d

Political debates

- $\boldsymbol{w}_d$: debate turn = document

Obamacare fight reenergizes tea party movement

- y_d: ideology of speaker



Data

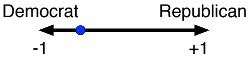
- A collection of documents $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_D$
- Each document d has an associated **response variable** y_d

Political debates

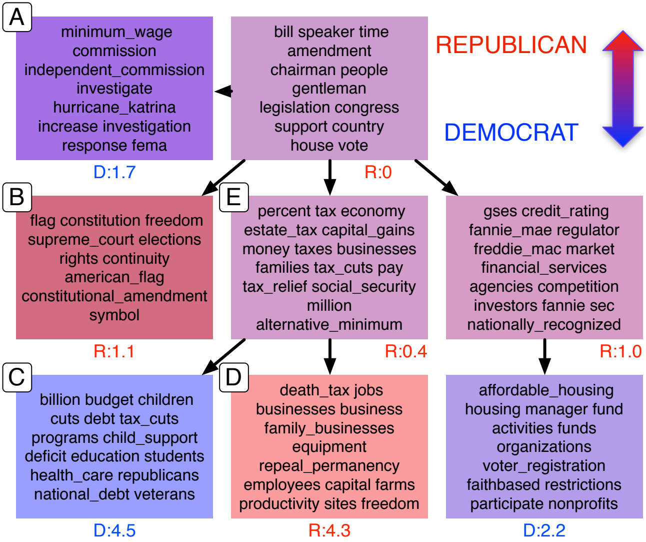
- $\boldsymbol{w}_d$: debate turn = document

Obama Blames Boehner for 'Reckless Republican Shutdown'

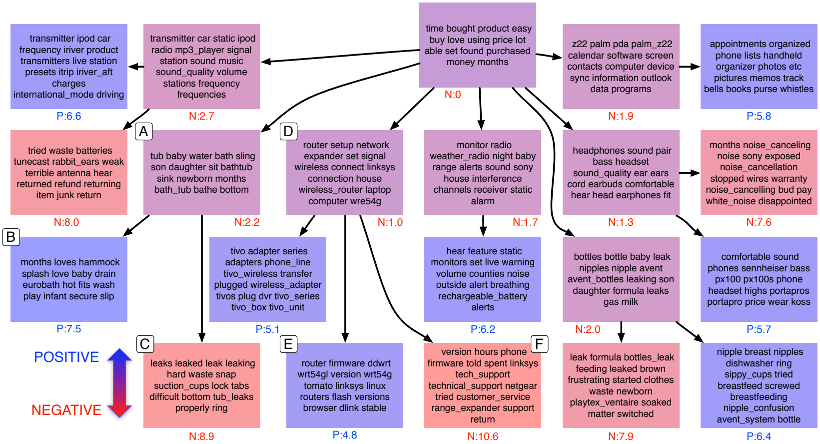
- y_d: ideology of speaker



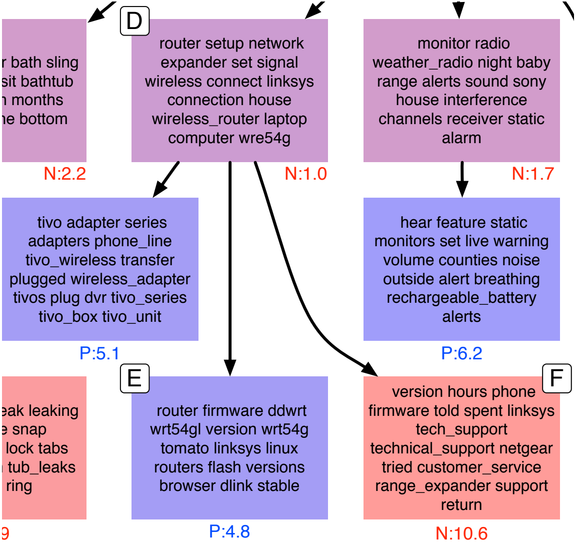
Qualitative results



Qualitative results



Qualitative results

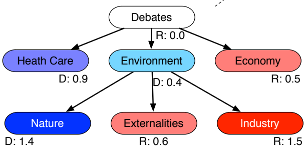


SHLDA: Generating response variable

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

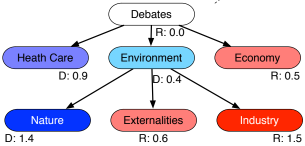
SHLDA: Generating response variable

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



SHLDA: Generating response variable

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

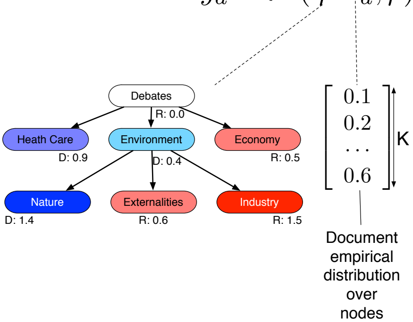


$$\begin{bmatrix} 0.1 \\ 0.2 \\ \cdots \\ 0.6 \end{bmatrix} K$$

Document empirical distribution over nodes

SHLDA: Generating response variable

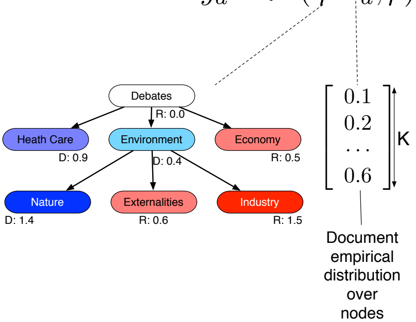
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho)$$



- Some words have **context-specific** contributions (topics)
  - "unpredictable": good for books, bad for steering

SHLDA: Generating response variable

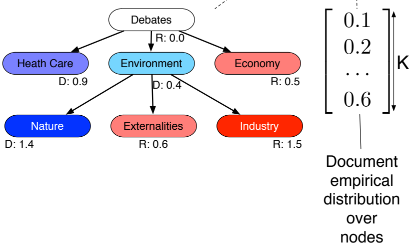
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d, \rho)$$



- 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

SHLDA: Generating response variable

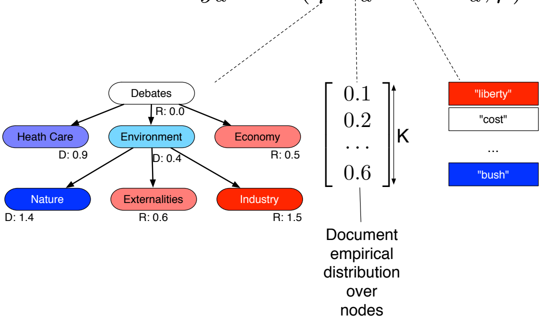
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

SHLDA: Generating response variable

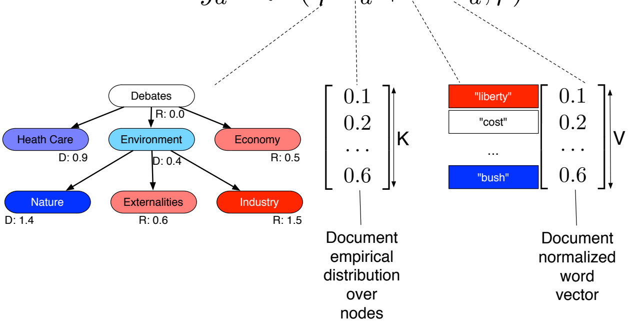
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

SHLDA: Generating response variable

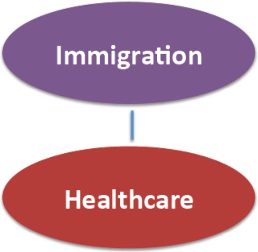
$$y_d \sim \mathcal{N}(\eta^T \bar{z}_d + \tau^T \bar{w}_d, \rho)$$



- 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

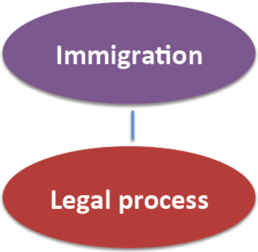
Swine flu surfaces at Texas-Mexico border among illega...
Washington Times - 10 hours ago
The first case of swine flu has been found among the scores of **illegal children** who have been crossing into America at the Texas-Mexico ...

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.

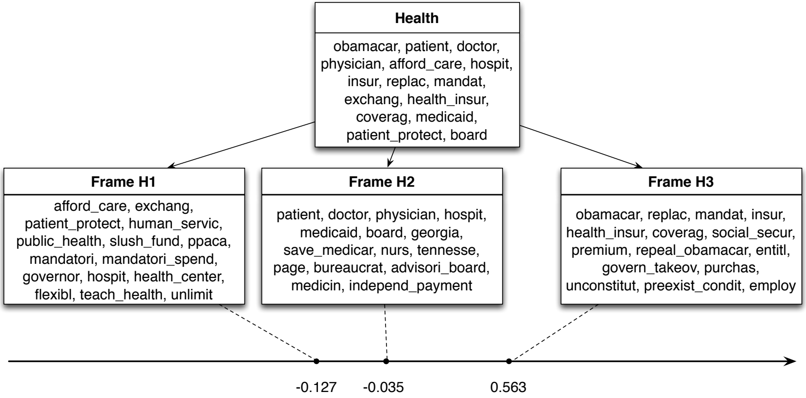


"Some of the children who have come to this country may not have a valid legal basis to remain, but some will. Yet**, it is virtually impossible for a child to assert a valid claim under immigration law in the absence of legal representation.** ... It is a fantasy to believe that unrepresented children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

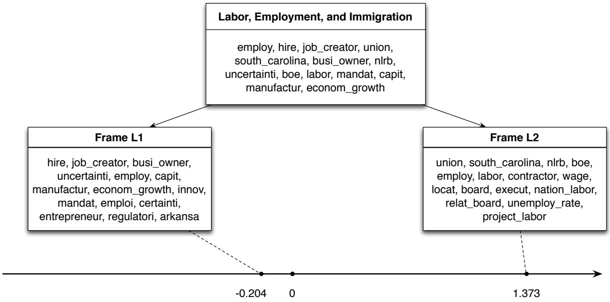
--Rep. Hakeem Jeffries, D-N.Y.



Issue-specific Framing: Health



Issue-specific Framing: Labor & Employment



Baumgartner, F. R. (2001).
Political agendas.
In Smelser, N. J. and Baltes, P. B., editors, *International Encyclopedia of Social and Behavioral Sciences: Political Science*, pages 288–90. New York: Elsevier Science and Oxford: Pergamon.

Bonica, A. (2013).
Ideology and interests in the political marketplace.
American Journal of Political Science, 57(2):294–311.

Carmines, E. G. and D'Amico, N. J. (2015).
The new look in political ideology research.
Annual Review of Political Science, 18(4).

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

Entman, R. M. (1993).
Framing: Toward clarification of a fractured paradigm.
Journal of Communication, 43(4):51–58.

Gerrish, S. and Blei, D. M. (2012).
How they vote: Issue-adjusted models of legislative behavior.
In *Proceedings of Advances in Neural Information Processing Systems*, pages 2753–2761.

Heckman, J. J. and Jr., J. M. S. (1997).
Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators.
The RAND Journal of Economics, 28:142–189.

References II

Jackman, S. (2001).
Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking.
Political Analysis, 9(3):227–241.

Lauderdale, B. E. and Clark, T. S. (2014).
Scaling politically meaningful dimensions using texts and votes.
American Journal of Political Science, 58(3):754–771.

McCombs, M. (2004).
Setting the agenda: The mass media and public opinion.
John Wiley & Sons.

Poole, K. T. and Rosenthal, H. (1985).
A spatial model for legislative roll call analysis.
American Journal of Political Science, pages 357–384.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010).
How to analyze political attention with minimal assumptions and costs.
American Journal of Political Science, 54(1):209–228.

Sim, Y., Routledge, B., and Smith, N. A. (2015).
The utility of text: The case of Amicus briefs and the Supreme Court.
In *Association for the Advancement of Artificial Intelligence*.