Task

- Given a congressman and the text of a bill, can we predict how that congressman will vote on the bill?
- Provides method for quantifying relationships between congresspeople and bills, topics, and ideas.

Ideal Vectors

- Most previous work represented congresspeople as ideal points.
- Assumed all legislators and bills are single points in one-dimensional 'political space'.
- First prior attempt at prediction task made by Gerrish and Blei (2011).
- They developed ideal point topic model, integrates topic model similar to LDA for bill text with ideal point model for congresspeople. Used variational inference to approximate posterior distribution of topics.
- Our model represents legislators as ideal vectors in higher-dimensional space.
- Ideal vectors are easy-to-train, multidimensional representation of legislator ideology.

Acknowledgements

We would like to thank the Harvard Natural Language Processing Group for its support, as well as Sam Wiseman and Saketh Rama for useful guidance.

References


Model

- Simple bilinear model that uses low-dimensional embeddings to model each word in our dictionary and each congressman.
- Models the probability of congressman \( c \) voting “yes” on a bill containing words \( B \):

\[
p(y = \text{yea} | B, c) = \sigma\left(\mathbf{W} \left( \sum_{w \in B} \mathbf{e}_w / |B| \right) + \mathbf{b} \cdot \mathbf{v}_c\right)
\]

- Bills are represented with word embeddings \( \mathbf{e}_w \in \mathbb{R}^{d_{emb}} \) for word \( w \) to capture multivariate relationships between words and their meanings.
- Word embeddings are initialized with GLoVe.
- Congresspeople are represented by ideal vectors, \( \mathbf{v}_c \in \mathbb{R}^{d_{emb}} \) for congressman \( c \) (with \( d_{emb} = 10 \)).
- We train \( \mathbf{W} \in \mathbb{R}^{k \times d_{emb}} \) and bias \( \mathbf{b} \in \mathbb{R}^k \).

Data

- Derived from the GovTrack database.
- Contains all votes on the full text (not amendments) of bills or resolutions.
- Uses data from the 106th to 111th Congress.
- Only contains yes-or-no votes, omitting abstentions and “present” votes (in accordance with previous work).
- Consists of 4067 bills and over a million unique yes-or-no votes.

Analysis

- PCA projection of ideal vectors for both houses of 111th Congress.
- Republicans red, Democrats who voted for Affordable Care Act (ACA) blue, Democrats who voted against ACA yellow, independents green.
- Model learns how (majority) Democrats much more unified than (minority) Republicans.
- Model also learns how conservative Democrats (who vote against ACA) closer to Republicans than other Democrats are.
- Relative favorability of congresspeople towards ‘Enterprise’ versus ‘Science’ in 110th Congress.
- Coordinates are sigmoids of dot products of ideal vectors with normalized word vectors.
- GOP red, Democrats blue, independents green.
- Model learns how both parties broadly support science, but Republicans are more pro-business.
- Model learns stances of individuals: Ron Paul (Libertarian), Kevin McCarthy (mainstream Republican), Jeff Flake (budget hawk).

Lexical Properties

- Democrats economic veterans political exchange
- Republicans state opportunities budget promote

Final Results

- Top five words by cosine similarity for each party in the 110th Congress with stop words removed.
- Democratic words are mostly words of budget and government as Democrats were majority party.
- Republican words mostly emphasize Republican themes and values.

<table>
<thead>
<tr>
<th>Congress</th>
<th>YEA</th>
<th>GB</th>
<th>IDP</th>
<th>EMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td>83.0</td>
<td>-</td>
<td>79.5</td>
<td>84.9</td>
</tr>
<tr>
<td>107</td>
<td>85.9</td>
<td>-</td>
<td>85.8</td>
<td>89.7</td>
</tr>
<tr>
<td>108</td>
<td>87.1</td>
<td>-</td>
<td>85.0</td>
<td>91.9</td>
</tr>
<tr>
<td>109</td>
<td>83.5</td>
<td>-</td>
<td>81.5</td>
<td>88.4</td>
</tr>
<tr>
<td>110</td>
<td>82.7</td>
<td>-</td>
<td>80.8</td>
<td>92.1</td>
</tr>
<tr>
<td>111</td>
<td>86.0</td>
<td>-</td>
<td>85.7</td>
<td>93.4</td>
</tr>
<tr>
<td>Avg</td>
<td>84.5</td>
<td>89</td>
<td>83.1</td>
<td>90.6</td>
</tr>
</tbody>
</table>

Conclusion

- We developed a novel model for predicting congressional roll-call votes from bill text.
- Our model outperforms any previous model while being extremely simple.
- We introduce ideal vectors as a fast, simple, multidimensional alternative to ideal points.

Acknowledgements

We would like to thank the Harvard Natural Language Processing Group for its support, as well as Sam Wiseman and Saketh Rama for useful guidance.

References