Roll Call Vote Prediction with Knowledge Augmented Models

Pallavi Patil[♦] **Kriti Myer**[♦] **Ronak Zala**[♦] **Arpit Singh**[♦]
Sheshera Mysore[♦] **Andrew McCallum**[♦] **Adrian Benton**[♠] **Amanda Stent**[♠]
[♦]College of Information and Computer Sciences
University of Massachusetts Amherst
[♠]Bloomberg LP
{ppatil,kmyer,rzala,arpitsingh,smysore,mccallum}@cs.umass.edu
{abenton10,astent}@bloomberg.net

Abstract

The official voting records of United States congresspeople are preserved as roll call votes. Prediction of voting behavior of politicians for whom no voting record exists, such as individuals running for office, is important for forecasting key political decisions. Prior work has relied on past votes cast to predict future votes, and thus fails to predict voting patterns for politicians without voting records. We address this by augmenting a prior state of the art model with multiple sources of external knowledge so as to enable prediction on unseen politicians. The sources of knowledge we use are news text and Freebase, a manually curated knowledge base. We propose augmentations based on unigram features for news text, and a knowledge base embedding method followed by a neural network composition for relations from Freebase. Empirical evaluation of these approaches indicate that the proposed models outperform the prior system for politicians with complete historical voting records by 1.0% point of accuracy (8.7% error reduction) and for politicians without voting records by 33.4% points of accuracy (66.7% error reduction). We also show that the knowledge base augmented approach outperforms the news text augmented approach by 4.2% points of accuracy.

1 Introduction

Roll call votes are official records of how politicians vote on bills (potential laws) in the United States House of Representatives and Senate. Reliable prediction of these votes, using historical voting records and the text of bills, can be used to forecast political decisions on key issues, which can be informative for the electorate and other political stakeholders. Prior work has used politicians' voting records as a means to study their ideological stances (Poole and Rosenthal, 1985; Clinton et al., 2004), as well as roll call votes combined with the text of the corresponding bills to predict votes on newly drafted bills (Gerrish and Blei, 2012; Kraft et al., 2016; Kornilova et al., 2018). However, these approaches fail to make good predictions for the votes of politicians whose records are not established, such as new candidates for office – a time when this information can be most useful for the electorate. We hypothesize that additional sources of knowledge about new politicians can predict their future votes.

In this work we explore two sources of additional knowledge about politicians to better predict roll call votes: news article text about the politicians, and Freebase (Bollacker et al., 2008), which is a manually curated knowledge base (KB). Relevant news articles may contain words that are indicative of a politician's stance with respect to specific issues. A KB such as Freebase is likely to contain rich information such as events a congressperson attends, people they are related to, and personal details such as schools they were educated at; this information may be correlated with a politician's stance on specific issues (Sunshine Hillygus, 2005; Duckitt and Sibley, 2010; Kraut and Lewis, 1975). Information in a KB is likely to be more restricted, but more reliable, than information extracted from news articles.

We integrate these sources of information into the embedding based prediction model proposed in Kraft et al. (2016). We experiment with two representations for news articles: as the mean of the embeddings of the words in the article, and as a bag of words. To represent information in KBs, we first capture the KB relations using Universal Schema (US) (Riedel et al., 2013), and then construct relation embeddings using a neural network.

We evaluate the proposed approaches on multiple sessions of Congress under two settings: (1) with only politicians that are observed at train-

Proceedings of the 23rd Conference on Computational Natural Language Learning, pages 574–581
Hong Kong, China, November 3-4, 2019.

ing time, which is similar to the setting of prior work (Kraft et al., 2016; Kornilova et al., 2018) and (2) where a subset of politicians' voting patterns are never observed at training time, representing the new candidate for office setting. We establish a new state-of-the-art under the evaluation framework used by most prior work (Setting 1). We also show that our approach outperforms a state-of-the-art model in our evaluation framework (Setting 2). Compared to the previous state-of-the-art model for roll call prediction, in Setting (1) our best approach gives an improvement in accuracy of 1.0% (an error reduction of 8.7%), and in Setting (2) our best approach gives an improvement in accuracy of 33.4% (an error reduction of 66.7%). Additionally, in Setting (2), augmentation via KB gives 4.2% more accurate predictions than augmentation from news text.

Code to reproduce our experiments can be found here: `https://github.com/ronakzala/universal-schema-bloomberg/`.

2 Models

All the models we explore take as input the voting record (denoted $\mathcal{V}$) for a politician p, the text of a bill b, and (optional) knowledge about the politician (denoted $\mathcal{K}$), and output a probability $\mathrm{P}(y = \texttt{Yea}|b, \mathcal{V}, \mathcal{K})$ that politician p will vote `Yea` on bill b. In this discussion, all features (voting record and knowledge augmentation) as well as labels (vote predictions) are specific to p, but for ease of reading we don't subscript to p.

We start with a baseline model that does not use any additional knowledge. We augment this model with knowledge from news articles mentioning p (denoted $\mathcal{N}$) and from parts of the Freebase KB relevant to p (denoted $\mathcal{F}$) (i.e. $\mathcal{N} \subset \mathcal{K}$ and $\mathcal{F} \subset \mathcal{K}$). In our augmented models, vector representations of this additional knowledge are denoted as $\mathbf{v}_{\mathcal{N}}$ and $\mathbf{v}_{\mathcal{F}}$. These vectors can be read as complementary representations of politician p, derived from how the politician appears in news articles and in curated sources of world knowledge, rather than by prior voting behavior. These vectors are used in conjunction with the politician's representation in terms of their historical voting record, denoted as $\mathbf{v}_{\mathcal{V}}$, which is learned as a model parameter. Now, we describe the forms of the baseline and augmented models.

2.1 Baseline Model

Our baseline is the model proposed by Kraft et al. (2016). This model represents a politician p by a vector $\mathbf{v}_{\mathcal{V}}$ and a bill b using a bag of words of the 1,000 most frequent unigrams across all bills, after excluding stopwords and single-character tokens. The model is defined as follows:

$$\mathbf{v}_b = \sum_{w \in b} \mathbf{e}_w / |b|$$
$$\mathbf{Z}_b = \mathbf{W}_{bill} \cdot \mathbf{v}_b + \mathbf{d}_{bill}$$
$$\mathrm{P}(y = \texttt{yea}|b, \mathcal{V}) = \sigma(\mathbf{Z}_b \cdot \mathbf{v}_{\mathcal{V}})$$

Here, $e_w \in \mathbb{R}^{d_{word}}$ is initialized to pre-trained GloVe word embeddings[1] and finetuned as a model parameter. $\mathbf{v}_{\mathcal{V}} \in \mathbb{R}^{d_{emb}}$ is the embedding for politician p and is initialized uniformly at random in $[-10^{-2}, 10^{-2}]$. Bill b is represented as $\mathbf{v}_b$, the average of word embeddings $\mathbf{e}_w$, and then transformed into the same space as $\mathbf{v}_{\mathcal{V}}$ via weight matrix $W_{bill} \in \mathbb{R}^{d_{emb} \times d_{word}}$ and bias vector $d_{bill} \in \mathbb{R}^{d_{emb}}$, i.e. $\mathbf{Z}_b \in \mathbb{R}^{d_{emb}}$.

2.2 News Text Augmented Model

We incorporate knowledge about politician p from relevant news articles in the form of unigram features extracted from the set of news articles that mention p, denoted $\mathcal{N}$. This model represents each article $a_j \in \mathcal{N}$ as a bag of words of the most frequent 2,000 unigrams across all articles for all politicians, after excluding stopwords and single-character tokens. The news augmented model is formulated as:

$$\mathbf{v}_b = \sum_{w \in b} \mathbf{e}_w / |b|$$
$$\mathbf{Z}_b = \mathbf{W}_{bill} \cdot \mathbf{v}_b + \mathbf{d}_{bill}$$
$$\mathbf{Z}_{\mathcal{N}} = \mathbf{W}_{news} \cdot \mathbf{v}_{\mathcal{N}} + \mathbf{d}_{news}$$
$$\mathrm{P}(y = \texttt{yea}|b, \mathcal{V}, \mathcal{N}) = \sigma(\mathbf{Z}_b \cdot [\mathbf{v}_{\mathcal{V}} + \mathbf{Z}_{\mathcal{N}}])$$

Most of the notation above is same as the one for the baseline model in §2.1. Here, $\mathbf{v}_{\mathcal{N}}$ is a vector representing the knowledge about p contained in N. $\mathbf{Z}_{\mathcal{N}} \in \mathbb{R}^{d_{emb}}$ represents $\mathbf{v}_{\mathcal{N}}$ transformed into the space of $\mathbf{v}_{\mathcal{V}}$ via weight matrix $\mathbf{W}_{news}$ and bias vector d_{news}. We experiment with two variations for computing $\mathbf{v}_{\mathcal{N}}$. First we compute $\mathbf{v}_{\mathcal{N}}$ as the mean GloVe vectors of all unigrams in $\mathcal{N}$

[1] `http://nlp.stanford.edu/data/glove.6B.zip`

(model denoted by NwGL):

$$\mathbf{v}_{\mathcal{N}}^{\text{GloVe}} = \frac{1}{\sum_{a \in \mathcal{N}} |a|} \sum_{a \in \mathcal{N}} \sum_{w \in a} \mathbf{e}_w$$

where $\mathbf{e}_w$ represents the GloVe vector for word w in article a. We also consider a variant where each article $a \in \mathcal{N}$ is represented as a vector of relative frequencies of the words in a (model denoted by NwFR):

$$\mathbf{v}_{\mathcal{N}}^{\text{FREQ}} = \sum_{a \in \mathcal{N}} \mathbf{f}_a$$

where $\mathbf{f}_a$ is a vector of unigram relative frequencies in article $a \in \mathcal{N}$.

2.3 Knowledge Base Augmented Model

Building on the baseline model, the knowledge base augmented model (denoted by KBUS) represents the contextual information for politician p as a vector $\mathbf{v}_{\mathcal{F}}$. The KB augmented model is formulated as:

$$\mathbf{v}_b = \sum_{w \in b} \mathbf{e}_w / |b|$$

$$\mathbf{Z}_b = \mathbf{W}_{bill} \cdot \mathbf{v}_b + \mathbf{d}_{bill}$$

$$\mathrm{P}(y = \texttt{yea} | b, \mathcal{V}, \mathcal{F}) = \sigma(\mathbf{Z}_b \cdot [\mathbf{v}_{\mathcal{V}} + \mathbf{Z}_{\mathcal{F}}])$$

Most of the notation above remains the same as the baseline model in §2.1. $\mathbf{Z}_{\mathcal{F}}$ is created from an embedded subgraph of the Freebase KB. Freebase consists of relation triples of the form (e_1, r, e_2), where e_1, and e_2 denote entities (e.g., `Barack Obama`, `Columbia University`) and r denotes a relation (e.g., `graduate_of`). These relation triples are embedded in a vector space by Universal Schema (Riedel et al., 2013), giving vector representations $\mathbf{v}_{e_1}$, $\mathbf{v}_{e_2}$, and $\mathbf{v}_r$ for the elements of the triple, e_1, e_2, and r. Universal Schema embeddings for entities and relations are trained for the KB completion task, to maximize the probability of triples existing in the KB parameterized by $\mathrm{P}(r, e_1, e_2) = \sigma(\mathbf{v}_r^T [\mathbf{v}_{e_1}; \mathbf{v}_{e_2}])$.

The KB knowledge vector $\mathbf{Z}_{\mathcal{F}}$ is derived from the Universal Schema entity embeddings. For each politician p, we consider the subset of relations in which the politician is either e_1 or e_2, denoted $\mathcal{F}$. We link p to a KB entity in Freebase by exact textual match on p's name. Let r_i be a relation in a triple in $\mathcal{F}$ and o_i be the entity that is *not* politician p in that triple; we compute $\mathbf{Z}_{\mathcal{F}}$ as:

$$\mathbf{Z}_{\mathcal{F}} = \sum_{i \in \mathcal{F}} \mathbf{t}_i \cdot \mathrm{P}(r_i, o_i, p)$$

$$\mathbf{t}_i = \mathrm{FFN}([\mathbf{v}_{r_i}; \mathbf{v}_{o_i}])$$

Here $[;]$ denotes concatenation, v_{r_i} the Universal Schema embedding for relation r_i, and v_{o_i} the Universal Schema embedding for o_i. $\mathbf{t}_i$ is the embedding of the ith triple and is computed by passing v_{r_i} and v_{o_i} through a shallow feed forward network (FFN) of the form: $tanh(\mathbf{W}\mathbf{x} + \mathbf{b})$. The parameters of this FFN are learned in the course of model training, and the dimensionality of $\mathbf{t}_i$ is a hyperparameter. The final representation for the contextual knowledge from the KB for politician p is therefore a sum of the Freebase triple Universal Schema embeddings weighted by the probability that p would participate in each triple.

3 Experiment Description

We evaluate the proposed knowledge augmented models under two evaluation protocols (§3.3): one where all politicians' voting records are observed in the training data (Setting 1) and another in which some politicians are unobserved at training time, representing the new candidate for office setting (Setting 2). We also evaluate the performance of simple baseline methods and previous state-of-the-art methods (§3.2) in both settings. Finally, we perform a close comparison of baseline methods against the proposed methods and highlight individual politicians for which the knowledge augmented models better predict voting behavior than the baselines (§4.0.4).

3.1 Datasets

Our data comes from three sources: roll call voting records from Congress Sessions 106-109, news articles from the Concretely Annotated New York Times Corpus (Ferraro et al., 2014), and the KB from Freebase. Here we briefly describe each source and major preprocessing decisions.

3.1.1 Roll Call Votes

The roll call votes dataset was compiled to resemble the one used by Kraft et al. (2016). The dataset covers the 106th-109th Congress (1999-2006), where each session spans 2 years, and was created by querying GovTrack[2] using pub-

[2]Online Roll Call Votes: https://www.govtrack.us/

licly available tools[3]. Bill texts, bill metadata, and lists of roll call votes were all queried separately and matched using bill IDs. The resulting dataset contains voting records for 709 unique politicians across sessions and almost 1 million politician-bill-vote examples. In order to facilitate future research on this data we make our tools and raw data available to the community, a component absent from prior work[4].

3.1.2 New York Times Corpus

We use the The Concretely Annotated NYT Corpus (Ferraro et al., 2014) as a source of knowledge contained in news articles. This corpus contains approximately 1.8 million articles for the period 1987-2007, which includes the sessions of Congress for which we report experimental results. The corpus is automatically annotated using the CoreNLP package with a host of annotations such as part of speech and named entity tags.

To obtain the set of politician-relevant articles for the news article augmented model (§2.2), we first construct a list of politician names from our roll call vote data and Wikipedia aliases for these names. We identify candidate names in each article by extracting all spans in the article annotated with the `Person` entity type tag. We then look for exact string matches to the list of politician names in the candidate names extracted from each article. By this means, we identify a total of 50,800 relevant articles. Each article is represented as a bag of words from the 2,000 most frequent words across all articles after dropping stop words and single character words. When training and evaluating our models, we conservatively include only the subset of relevant articles published before the congressional session from which the bill comes; this way, information from news text necessarily predates the congressional votes and does not inform about the model about voting outcomes.

3.1.3 Freebase

We use Freebase as a source of KB knowledge. Freebase is a large, structured knowledge base that consists of relation triples created by human contributors. It contains about 46 million unique entities and 332 million relation triples.

In extracting relations relevant to politicians in our roll call vote data, we first filter to only relations that mention a politician explicitly as one of the entities. We add to this set relations one hop away from the politicians in the knowledge base (the politicians do not directly participate in this set of one hop away relations). The combination of direct and one hop away relations gives about 800,000 relation triples. This subset of Freebase is embedded with Universal Schema, following which only the direct politician relations, along with the scores learned by Universal Schema for these relations, are used in our KB augmented model (§2.3). Although Freebase relations are not temporally marked, Freebase does not contain politicians' voting records, so vote information cannot "bleed into" our models from Freebase.

3.2 Baselines

We compare the performance of the proposed models against several baseline and previous state-of-the-art models. These are:

KJR16: The model from Kraft et al. (2016). The numbers we report for KJR16 are based on our re-implementation of the original model. We also apply slightly different preprocessing decisions to the text of bills. For example, we drop stop words and single-character tokens, unlike the original paper.

MAJ: This baseline model predicts the majority class (`Yea`) for all votes on all bills.

PARTYM: This baseline predicts that a congressperson will vote for a bill in the direction of their party's majority vote for that bill. A congressperson can have a party affiliation of Republican, Democrat, or Independent. A politician's party is generally a very strong predictor of their voting behavior. We compare against this baseline specifically to measure how much gain contextual knowledge about a politician can bring over just predicting that a congressperson will toe the party line. This baseline operates in a slightly unrealistic setting, since the party majority vote is determined *after* all the votes have been cast for a given bill.

3.3 Experimental Setup

For all experiments, our models are trained and evaluated on each Congress session separately. This setup resembles that of most prior work. One exception is Kornilova et al. (2018), who evaluate

[3] Tools to query online congressional data sources: https://github.com/unitedstates/congress

[4] https://github.com/ronakzala/universal-schema-bloomberg/

Session	Majority Class	Accuracy (%)		Precision (%)		Recall (%)		F1 (%)	
		KJR16	BP	KJR16	BP	KJR16	BP	KJR16	BP
106	83.03	85.90	**86.49**	**90.46**	89.14	92.78	**95.32**	91.60	**92.13**
107	85.86	91.10	**91.63**	91.65	**92.57**	98.06	**98.51**	95.06	**95.44**
108	87.10	90.64	**91.68**	91.88	**93.64**	**97.27**	97.15	94.38	**95.36**
109	83.48	86.24	**88.07**	91.85	**92.28**	91.26	**93.05**	91.05	**92.66**
Average	84.87	88.47	**89.47**	91.46	**91.91**	94.84	**96.00**	93.02	**93.89**

Table 1: Performance of KJR16 (our reimplementation of Kraft et al. (2016)) compared to our Best Proposed (BP) model (News augmented-Glove, NWGL), in evaluation Setting (1).

on the next Congress session; even in this case, they restricted their evaluation sets to only those politicians that were observed at training time.

We use a train/dev/test split of 60%/20%/20%; every session of Congress contains about 250,000 politician-bill-vote examples given that each congress contains between 500-650 bills, and approximately 550 politicians. All of the reported numbers are averaged over four random restarts of the models.

As noted earlier, we evaluate the models under two settings: (Setting 1) in this setting all politicians are present in both training and test data (this is the setting used in all prior work); and (Setting 2) in this setting, votes from 5% of the politicians, chosen at random for each session of Congress, are removed from the training set[5], resulting in a reduction of around 7,000 politician-bill-vote examples from each session of Congress. All of these politicians are still present in the test set.

3.4 Model Hyperparameters and Training

All the models we train (KJR16, NWGL, NWFR, KBUS) use a bill embedding ($\mathbf{v}_b$) of size 50, and a per-politician embedding $\mathbf{v}_{\mathcal{V}}$ of size 10. The NWGL model additionally uses a mean GloVe vector ($\mathbf{v}_{\mathcal{N}}^{\text{GloVe}}$) of size 50, whereas the NWFR model uses a word frequency vector ($\mathbf{v}_{\mathcal{N}}^{\text{FREQ}}$) of size 2,000. The KBUS model has an entity embedding ($\mathbf{v}_o$) of size 25, and relation embedding ($\mathbf{v}_r$) of size 25, resulting in a KB knowledge vector ($\mathbf{Z}_{\mathcal{F}}$) of size 10 using the FFN architecture: $\{50, 10\}$. These settings were chosen without exploration; further hyperparameter tuning may result in different model performance.

All models were trained using vanilla SGD with a learning rate of 0.1 for up to 20 epochs[6]. We trained with early stopping based on the accuracy on the development set of politician-bill pairs.

For all models, we report accuracy, precision, recall, and F averaging over all vote predictions (i.e. we micro-average).

[5] On average, 2-10% of politicians are newcomers during every session, making our removal of 5% politicians realistic.

[6] No momentum or minibatching was used.

4 Results and Analysis

Our experiments attempt to answer several questions: (1) Setting 1: Does politician-related knowledge augmentation improve predictive performance when voting records of all politicians are observed (§4.0.1)? (2) Setting 2: Does politician-related knowledge augmentation improve performance when voting records of some politicians are not observed (§4.0.2)? (3) Does knowledge augmentation from a manually curated KB improve model performance compared to knowledge augmentation using unstructured text (§4.0.3)? (4) Finally, for which politicians are our knowledge augmented models more effective than predicting a party majority (PARTYM) vote (§4.0.4)?

4.0.1 Setting 1: All Voting Records Observed

Tables 1 and 2a display the performance of our proposed models when the voting patterns of all politicians are observed. Our best proposed model NWGL outperforms KJR16 in all sessions of Congress according to most metrics (Table 1). NWGL outperforms KJR16 by 1% point of accuracy on average, an error reduction of 8.7%. Knowledge augmentation in any form (NWGL, NWFR and KBUS) gives small improvements in roll call vote prediction over KJR16 (Table 2a).

4.0.2 Setting 2: Some Voting Records Absent

Table 2b displays results in the setting where some politicians' voting data was removed from the training set, representing the new candidate for office setting. For this setting, KJR16 makes random predictions. By contrast, all our models are able to

Model	Accuracy (%)	F1
MAJ	*83.82*	*91.19*
PARTYM	*83.56*	*90.54*
KJR16	88.47	93.17
NWGL	**89.47**	**93.89**
NWFR	89.33	93.78
KBUS	89.19	93.68

(a) Setting 1: All politicians in the test set are present in the training set.

	All		Absent from train		Present in train	
Model	Accuracy (%)	F1 (%)	Accuracy (%)	F1 (%)	Accuracy (%)	F1 (%)
MAJ	*83.82*	*91.19*	*83.65*	*91.07*	*83.82*	*91.19*
PARTYM	*83.56*	*90.54*	*84.59*	*91.07*	*83.53*	*90.52*
KJR16	84.31	90.73	49.96	62.37	85.51	91.48
NWGL	85.24	91.23	78.40	86.79	85.48	91.38
NWFR	83.65	90.54	79.13	87.66	83.80	90.64
KBUS	**85.98**	**91.96**	**83.36**	**90.52**	**86.07**	**92.01**

(b) Setting 2: An arbitrary 5% of politicians in every session of Congress are absent from the training set.

Table 2: Performance of our proposed models: News augmented-Glove (NWGL), News augmented-Word Frequency (NWFR) and KB Augmented-US (KBUS). These are compared to baselines Majority Class (MAJ) and Majority Party (PARTYM), and to the model from Kraft et al. (2016) (KJR16) for both evaluation settings. Here we report performance averaged across all four sessions of Congress, and highlight the best model excluding MAJ and PARTYM, which are unrealistic solutions.

successfully predict many of the votes of unseen politicians. The KB augmented model, KBUS, outperforms KJR16 by 33.4% points of accuracy, an error reduction of 66.7%.

MAJ and PARTYM outperform the proposed models in Setting 2. However, PARTYM is based on knowledge of the future and so is impractical, and MAJ has no discriminative power for individual politicians.

4.0.3 Comparison of Augmented Models

The KB augmented model KBUS relies on structured contextual information present in relation triples while the news text augmented models NWGL and NWFR rely on unstructured representations of contextual information (average of unigram embeddings). Also, the KB is manually curated while the news augmented models rely on noisier automatically selected and processed sources of information.

In Setting 1 (the setting in which all politicians are present in the training data) all three augmented models perform similarly, with NWGL and KBUS slightly outperforming NWFR (Table 2a). However, in Setting 2 (where some politicians are unseen in the training data) KBUS clearly outperforms NWGL and NWFR, both overall (85.98% accuracy) and specifically on the unseen politicians (83.36% accuracy). We conclude that the structured nature of the additional knowledge provided to this model might allow for more effective use of contextual knowledge. This also suggests that the news text augmented model can be improved by improving the quality of news text features or learning a more clever weighting of unigram embeddings, for example by the CNN embedding approach used in Kornilova et al. (2018).

4.0.4 Comparison with PARTYM

Here we present a deeper analysis of specific politicians to highlight cases where rich contextual information aids prediction as compared to a model only relying on party affiliation.

We examine results for the 108th Congress, where our KB augmented model KBUS achieves an accuracy of 91.39%, while the PARTYM baseline has accuracy of only 85.26%. The congressperson embeddings learned by KBUS also appear to capture party affiliation (as demonstrated by the near-linear separability of parties evident in Figure 1), but they strictly outperform PARTYM for this session. The embeddings learned by

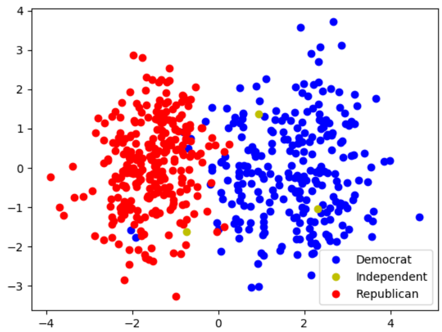

Figure 1: 2-dimensional PCA projected congressperson embeddings by the KBUS model for the 108th congress. Points are colored by party affiliation.

PARTYM primarily err under two circumstances: politicians who do not always vote the party line, and those who identify themselves as Independent. We examine a handful of individual politicians to highlight cases where the proposed models outperform the PARTYM baseline:

Politicians straying party lines : Joe Baca was a congressman from California (D-CA) who was a Democrat during the 108th Congress, but has since changed party twice, due to his Republican-leaning ideology. For Congressman Baca KBUS's accuracy is 90.9%, compared to PARTYM's accuracy of 79.09%. As another example, Jeff Flake (R-AZ) was at that time a congressman recognized as a traditional conservative, but voted with Democrats on issues such as immigration and employment non-discrimination. For Congresssman Flake, KBUS's accuracy is 83.92% while PARTYM's accuracy is 76.78%.

Independent candidates : Joseph Lieberman was an Independent senator from Connecticut (I-CT). For Senator Lieberman, KBUS's accuracy is 91.67% while PARTYM's accuracy is 83.33%. Votes cast by Independents cannot be reliably predicted by considering how other Independents would vote on the same bill.

These results further highlight the importance of a model that is able to combine voting history with rich contextual knowledge in order to accurately predict votes.

5 Related Work

Prior work on predicting roll call votes on bills is primarily focused on using Ideal Point Models, which assume that a politician's ideology and the ideology reflected in a bill lie along a single dimension. In Clinton et al. (2004), an Ideal Point model was trained over both politicians and bills/issues, and at inference time the similarity between politician and bill was used to determine how likely the politician was to vote for the bill. However, most politicians have distinct views on different issues, meaning that the views of one politician cannot be captured in a single dimension. Recent work has taken this into account. For example, Gerrish and Blei (2012) created a model in which each politician's ideal point is adjusted per issue, based on the bill text.

Kraft et al. (2016) used an embedding based model which jointly learns bill and politician embeddings, and can predict how a politician will vote on a bill. As demonstrated in this paper, this model works well when a politician's voting track record has already been established. However, it fails for politicians not in the training data, such as those who have never been elected to office or voted on bills relevant to the issues in the target bill. Although we did not implement any Ideal Point Models, they obviously share this weakness.

There have also been approaches that incorporate additional knowledge about politicians and bills to yield extra insight into politicians' voting patterns. Kornilova et al. (2018) enhanced the embeddings learned by their model by providing bill sponsorship information along with a CNN architecture for learning bill embeddings, while Nguyen et al. (2015) supplemented their model by taking into consideration the type of language legislators use. However, these models share the inability of earlier models to predict voting behavior for new politicians or political candidates. In fact, Kornilova et al. (2018) explicitly note that their model cannot handle unobserved politicians: they say, "During testing, we only include legislators present in the training data". The contributions of Kornilova et al. (2018), among them the CNN-based bill representation and the incorporation of bill metadata, are orthogonal to those in this paper. We leave the exploration of how to combine these two approaches to future work.

Our work builds on the idea of using additional knowledge about politicians to enhance vote prediction performance. The model described by Kraft et al. (2016) serves as our baseline model, and our proposed augmented models build on this baseline by supplementing it with additional knowledge about the politicians. Unlike Kornilova et al. (2018), the additional knowledge we inject is about politicians rather than bills. This allows our proposed models to generalize to politicians unseen at training time.

6 Conclusion and Future Work

In this paper we proposed methods for augmenting a state of the art model (Kraft et al., 2016) for roll call vote prediction with rich sources of additional knowledge to facilitate prediction in cases where the voting record for a politician is unavailable at training time. This is typically the case for new candidates for office or newly elected politicians. We demonstrate that our proposed models outperform a previous state-of-the-art model both when the voting record for all politicians in the test set is known and when the voting record of some politicians is not available at training time.

We propose several avenues for future research. First, researchers could explore richer representations for text, both bill text and news text - for example, CNNs or contextual language models like BERT (Devlin et al., 2019). Second, researchers could explore methods for knowledge augmentation using open information extraction or other ways of automatically constructing KBs, to reduce reliance on manually curated knowledge sources which are subject to bias and quickly become out of date. And third, models could be explored that can take into account interactions between politicians (e.g. changes in majority party, changes in politicians' affiliation, party movement leftward or rightward), between bills (e.g. bill combination or revision), and between politicians and bills across sessions of Congress (e.g. bill revision).

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*.

Joshua Clinton, Simon Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review*, 98(2):355–370.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

John Duckitt and Chris G. Sibley. 2010. Personality, ideology, prejudice, and politics: A dual-process motivational model. *Journal of Personality*, 78(6):1861–1894.

Francis Ferraro, Max Thomas, Matthew Gormley, Travis Wolfe, Craig Harman, and Benjamin. Van Durme. 2014. Concretely annotated corpora. Presented at the NIPS Workshop on Automated Knowledge Base Construction (AKBC).

Sean Gerrish and David M. Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems 25*.

Anastassia Kornilova, Daniel Argyle, and Vladimir Eidelman. 2018. Party matters: Enhancing legislative embeddings with author attributes for vote prediction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Peter Kraft, Hirsh Jain, and Alexander M. Rush. 2016. An embedding model for predicting roll-call votes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Robert E Kraut and Steven H Lewis. 1975. Alternate models of family influence on student political ideology. *Journal of Personality and Social Psychology*, 31(5):791.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to Republican legislators in the 112th Congress. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Keith T Poole and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science*, pages 357–384.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

D. Sunshine Hillygus. 2005. The missing link: Exploring the relationship between higher education and political engagement. *Political Behavior*, 27(1):25–47.