Align Voting Behavior with Public Statements for Legislator Representation Learning

Xinyi Mou[1], Zhongyu Wei[1,2]*, Lei Chen[1], Shangyi Ning[1],
Yancheng He[3], Changjian Jiang[4]*, Xuanjing Huang[5]

[1]School of Data Science, Fudan University, China
[2]Research Institute of Intelligent and Complex Systems, Fudan University, China
[3]Platform and Content Group, Tencent, China
[4]School of International Relations & Public Affairs, Fudan University, China
[5]School of Computer Science, Fudan University, China
{xymou20,zywei,chenl18,syning18}@fudan.edu.cn
collinhe@tencent.com, {Changjian,xjhuang}@fudan.edu.cn

Abstract

Ideology of legislators is typically estimated by ideal point models from historical records of votes. It represents legislators and legislation as points in a latent space and shows promising results for modeling voting behavior. However, it fails to capture more specific attitudes of legislators toward emerging issues and is unable to model newly-elected legislators without voting histories. In order to mitigate these two problems, we explore to incorporate both voting behavior and public statements on Twitter to jointly model legislators. In addition, we propose a novel task, namely hashtag usage prediction to model the ideology of legislators on Twitter. In practice, we construct a heterogeneous graph for the legislative context and use relational graph neural networks to learn the representation of legislators with the guidance of historical records of their voting and hashtag usage. Experiment results indicate that our model yields significant improvements for the task of roll call vote prediction. Further analysis further demonstrates that legislator representation we learned captures nuances in statements.

1 Introduction

Modeling the behavior of legislators is one of the most important topics of quantitative political science. Existing researches largely rely on roll call data, i.e. historical voting records, to estimate the political preference of legislators. The most widely used approach for roll call data analysis is ideal point model (Clinton et al., 2004) that represents legislators and legislation as points in a one-dimension latent space. Researchers enhance ideal point model by incorporating textual information of legislation (Gerrish and Blei, 2011; Gu
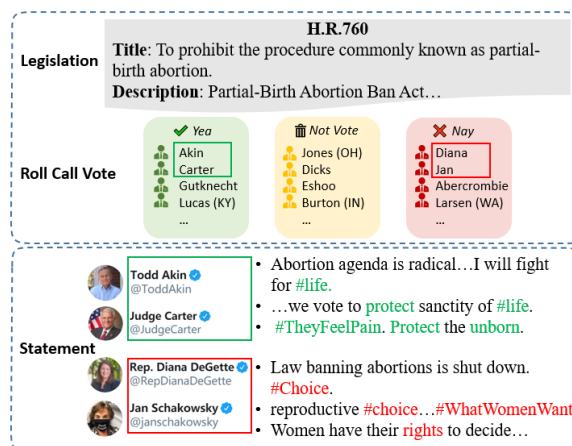

Figure 1: An illustration of correspondence of vote behavior and public statements on Twitter. Supporters of the abortion-banning legislation frequently mention the tag *life* while opponents focus on *choice*.

et al., 2014; Kraft et al., 2016) and report positive results for roll call vote prediction.

Although roll call data is the major resource for legislator behavior modeling, it has two limitations. Firstly, it fails to uncover detailed opinions of legislators towards legislative issues. Therefore, we have no clue about the motivation behind their voting. Secondly, it is unable to model the behavior of newly-elected legislators because their historical voting records are not available (i.e., cold-start problem). Meanwhile, researchers explore to use public statements to characterize the ideology of legislators with the guidance of framing theory (Entman, 1993; Chong and Druckman, 2007; Baumer et al., 2015; Vafa et al., 2020). Vafa et al. (2020) propose a text-based ideal point model to analyze tweets of legislators independent of roll call data. Experiment results show some correlations between distributions of ideal points learned

*Corresponding author.

1236

from legislative data and public statements. However, they treat the two resources separately and fail to uncover deep relationships of behavior between these two landscapes.

Figure 1 shows a legislative issue related to *prohibit partial-birth abortion*. It includes the title and description of the legislation, roll call vote records and public statements on Twitter of legislators. Based on the voting records, we know the stance of legislators. With the discussion on Twitter, we can further understand their opinions towards the topic. Supporters concentrate on *protecting the life* while opponents emphasize *rights of choice*. This motivates that bridging public statements on Twitter with roll call data can provide a full image of behavior patterns of legislators.

A closer look at the example (Figure 1) reveals that most tweets utilize hashtags to express ideas in short. Moreover, people with opposite stances choose different groups of hashtags, i.e., supporters use *#life* and *#TheyFeelPain* while opponents use *#Choice* and *#WhatWomenWant*. Further analysis on a large tweets dataset, where each tweet is processed by a python library TextBlob[1], shows that most hashtags are polarized with one sentiment (Figure 2a). Based on this observation and previous studies that reveal polarization of hashtags (Conover et al., 2011a; Garimella and Weber, 2017), we explore to utilize hashtags as a label to describe the preferences of legislators on public discussion and propose a novel task of hashtag usage prediction to characterize their ideology.

In this paper, we collect public statements of legislators on Twitter as an extension of roll call data for legislator representation learning. Our intuition is to combine roll call votes as hard labels and hashtags as soft labels to jointly model legislators. In practice, we build a heterogeneous graph to bridge the voting behavior and public statements of legislators. It consists of three kinds of nodes, legislators, legislation and hashtags in tweets. Subsequently, we employ a heterogeneous Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018) to simultaneously update the representation of different nodes. Two tasks are used for training, including roll call vote prediction and hashtag usage prediction to model the behavior of legislators on voting and on public statements respectively. The major contributions of this paper are threefold:
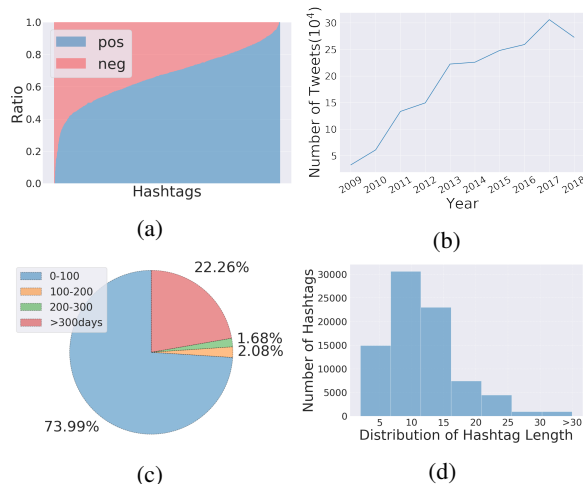
Figure 2: Statistics of Twitter Dataset. (a) sentiment distribution of hashtags in legislators' tweets. (b) number of tweets each year. (c) life span of hashtags. (d)distribution of length of hashtags

- To the best of our knowledge, this is the first study incorporating both voting behavior and public statements to jointly depict legislators. The proposed framework enables us to understand the preferences of legislators combining their behavior in legislative process and on public platforms.

- We propose to learn the representation of legislation and legislators using heterogeneous graph which can densify relations among legislators, thus mitigate the cold-start problem.

- We propose a novel task of hashtag usage prediction to characterize the preferences of legislators on public discussion and construct a dataset as the benchmark. Our dataset and code is available on Github [2].

2 Dataset and Tasks

The *Voteview* website (Lewis et al., 2021) provides a benchmark for the task of roll call vote prediction. It contains roll call votes history and keeps updating. Meanwhile, a dataset constructed by Yang et al. (2020) enables the public to take advantage of detailed description and sponsor information of legislation from 1993 to 2018. We extend these corpora with tweets published by legislators.

2.1 Twitter Dataset

Since Twitter became popular among legislators in the last decade, we reserve 1,198,758 roll call
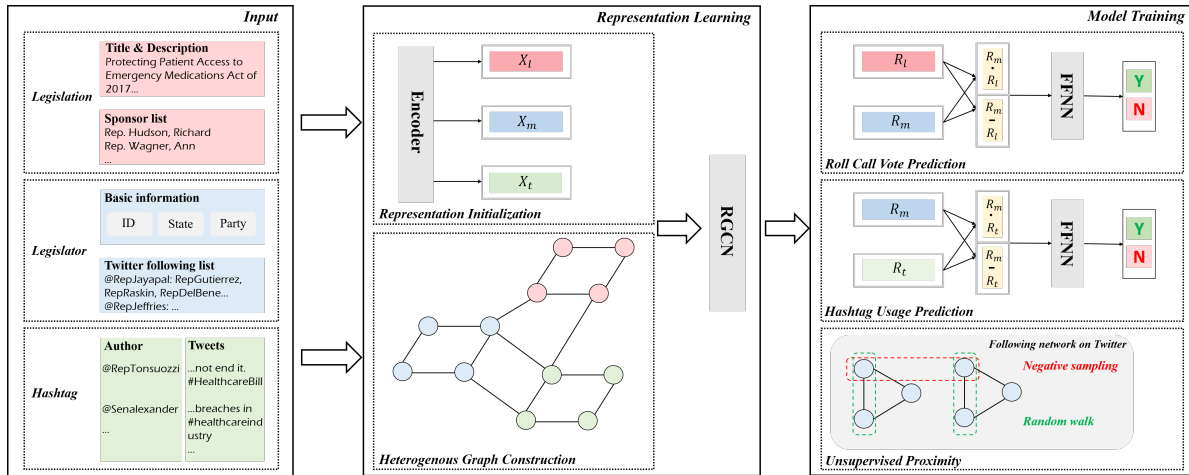
Figure 3: Proposed Framework.

records after 2009, involving 906 legislators and 3,210 pieces of legislation. For dataset construction, we first extract Twitter accounts of legislators from their homepages on the website of U.S. Congress [3]. For those who have not provided Twitter account, we manually search their names on Twitter, and identify their accounts by checking the verification information and biography. In this way, 735 accounts of legislators are included in our extended dataset. We crawl all tweets (before July 20th, 2020) for each legislator remained via twitterscraper [4]. In addition to this, we also collect their following list.

We show some statistics of the dataset in Figure 2. Figure 2b presents the distribution of the amount of tweets posted by year. It shows that legislators pay increasing attention to Twitter from year 2009 to 2017. Legislators post 3,071 tweets on average and 57.82% of legislators post more than 2,000 times. In terms of hashtag, a third of tweets contain at least one hashtag with 82,381 unique hashtags in total. Figure 2c indicates that most hashtags fade away within three months. Figure 2d shows the distribution of the length of hashtags, illustrating a hashtag usually consists of a few words. In order to reduce noise, we keep hashtags with length greater than 2 and frequency higher than 50. After that, 2,057 hashtags are reserved for graph construction.

To explore hashtag usage behavior, we construct 0-1 labels indicating whether a legislator has posted a specific hashtag or not. Considering some hashtags are not popular, we further remove those posted by less than 100 legislators, for hashtag

---

[3]www.congress.gov
[4]https://github.com/bisguzar/twitter-scraper

usage prediction. In this way, 194,040 labels are created.

2.2 Task Formulation

We introduce notations in this paper.

- $M = \{m_1, m_2, ...\}$ is the list of legislators, where each $m_i(i = 1, 2, ...)$ contains basic background information of legislators: member ID, state and party, accompanied with following list on Twitter.

- $L = \{l_1, l_2, ...\}$ is the list of legislation, where each $l_i(i = 1, 2, ..)$ contains its title and description, as well as sponsor information and voting results.

- $T = \{t_1, t_2, ...\}$ is the list of hashtags that have been mentioned by legislators on Twitter. Each of these hashtags contains information of related tweets and authors.

Note that each element (legislator, legislation or hashtag) is accompanied with the time when it appears in the context. We utilize these time markers to build our experimental environment to avoid future information leakage.

We use two tasks, i.e., roll call vote prediction and hashtag usage prediction to characterize the behavior of legislators in different landscapes, namely, Congress and Twitter. **(1) Roll call vote prediction**. This task aims to predict vote results of legislators towards legislation with stances of *yea* or *nay*. **(2) Hashtag usage prediction**. This task aims to predict whether a legislator will post a given hashtag or not.

1238

3 Proposed Framework

The overall framework we proposed is shown in Figure 3. We construct a heterogeneous graph with three kinds of nodes (legislation, legislator and hashtag) to cover the two landscapes of Congress and Twitter. On top of this graph, RGCN is applied to optimize the representation. This is achieved by a joint training of the two tasks of roll call vote prediction and hashtag usage prediction. In addition, we utilize an unsupervised following proximity loss to further optimize the representation.

3.1 Heterogeneous Graph Construction

The heterogeneous graph consists of three kinds of nodes and six types of relations with two categories (relations between homogeneous nodes and relations between heterogeneous nodes). We will introduce the structure of the graph in this subsection.

3.1.1 Initialization of Nodes

Legislator Nodes We follow Yang et al. (2020) to map each legislator to a continuous low-dimension vector, utilizing information of member ID, state and party. The legislator representation is $X_m = e_{ID} \oplus e_{Party} \oplus e_{State}$

Legislation Nodes For legislation, we pay attention to title and description and represent each legislation by sentence embedding generated by BERT (Devlin et al., 2019). Thus, the legislation representation is $X_l = BERT(title + description)$

Hashtag Nodes To represent a hashtag, we randomly choose K tweets with the tag and use BERT to get sentence embedding of each tweet text. After that, we take the average of these vectors, $X_t = Avg(BERT(tweet_i)) \; i = 1, 2, ...K$

3.1.2 Relations between Homogeneous Nodes

R1: Co-sponsorship of Legislators Each legislation is initialized by a sponsor and several co-sponsors. Previous study (Yang et al., 2020) has proved the effectiveness of modeling co-sponsorship in legislator representation learning. Obviously, more legislation two legislators have collaborated on means they are more alike ideologically. We follow this setup and regard the number of legislation two legislators have co-sponsored as weight of this relation to measure strength of the relationship between congressmen. In this way, a legislator network can be constructed and we obtain an adjacency matrix A, with each element a_{ij} representing the number of legislation m_i and m_j have co-sponsored.

R2: Similarity of Legislation Both topic models and embedding paradigms have been incorporated to model legislation in previous studies. However, the semantic relations among legislation have not been explicitly considered. We explore to better learn legislation representation by incorporating these semantic relationships. To achieve this goal, we construct a network of legislation, and use semantic similarity to link two legislation. Specifically, an adjacency matrix B is computed, with each element b_{ij} denoting the number of common words in texts of legislation l_i and l_j.

R3: Co-occurrence of Hashtags If two hashtags are mentioned together frequently, it's likely that they bear similar ideas, such as *#dreamact* and *#protectdreamers*. Therefore, we build a hashtag network, to help hashtag nodes learn from ones with similar ideology. An adjacency matrix C is constructed, with each element c_{ij} indicating the number of co-occurrence of hashtag t_i and t_j.

3.1.3 Relations between Heterogeneous Nodes

R4: Relation between Legislator and Legislation
In the legislative process, each legislation is initialized by multiple legislators. Karimi et al. (2019) have indicated that features of the bipartite network of legislators and bills are informative. Therefore, we use such sponsorship relation to connect nodes of legislator and legislation. An adjacency matrix D is constructed, with each element d_{ij} meaning whether legislator m_i has sponsored legislation l_j.

$$d_{ij} = \begin{cases} 1 & if \; m_i \; has \; sponsored \; l_j \\ 0 & otherwise \end{cases}$$

R5: Relation between Legislator and Hashtag
Legislators choose hashtags to use when they publish tweets. Therefore, we define an adjacency matrix F to measure preferences of legislators to hashtags. Each element f_{ij} is computed as the times legislator m_i has mentioned hashtag t_j.

R6: Relation between Legislation and Hashtag
Legislation might discuss similar topics with hashtags used in tweets. We therefore align legislation with hashtags by computing the semantic similarity based on their textual information. To achieve this, an adjacency matrix G is constructed, with each element g_{ij} representing the number of common

words in the text of legislation l_i and tweets with hashtag t_j.

3.2 Relational Graph Convolutional Network

After initializing representation of legislator, legislation and hashtag, we feed them into Relational Graph Convolutional Network(RGCN) (Schlichtkrull et al., 2018) to update their representation based on the context. Graph convolutional networks (GCNs) (Kipf and Welling, 2017) provide an efficient way to perform message propagation and aggregation. In the propagation phase, nodes send signals to their neighbors while in the aggregation phase, each node sums up messages from its neighbors and updates its representation. When there are only one type of relations, the layer-wise rule of GCNs is:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) \qquad (1)$$

where $H^{(l)}$ is hidden representation of lth layer, $\hat{A}$ represents the adjusted adjacency matrix and $W^{(l)}$ is weight matrix shared by all edges in layer l, $\sigma(\cdot)$represents the activation function. For each node i with neighbors $\mathcal{N}_i$, the update rule can be described as:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \frac{1}{c_i} W^{(l)} h_j^{(l)}\right) \qquad (2)$$

where c_i represents the normalization item, which is often set to $|\mathcal{N}_i|$ when each neighbor has equal importance.

RGCNs generalize GCNs to deal with relations of different types. RGCNs utilize different weight matrixes and normalization factors for different relation types. Thus, the hidden representation for each node i in layer $(l+1)$ can be computed as:

$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right) \qquad (3)$$

where $\mathcal{R}$ is the set of relation types, and $\mathcal{N}_i^r$ is the set of neighbors of node i connected by relation type r. Since each neighbor has different degrees of importance in our graph, we compute the normalization factor $c_{i,r}$ according to weights of relations we have obtained, instead of using $c_{i,r} = |\mathcal{N}_i^r|$. We apply 2-layer RGCNs to capture 2nd order relations between nodes empirically. After convolution, we get representations of legislator, legislation and hashtag, denoted as R_m, R_l and R_t.

3.3 Model Training

We utilize two tasks, namely roll call vote prediction and hashtag usage prediction to train our model. In addition, we introduce a following proximity loss to further measure relationships of legislators based on their social networks.

3.3.1 Roll Call Vote Prediction

Given representation of legislators and legislation, the roll call vote prediction comes out to be a classification task. We conduct element-wise product and element-wise difference of embeddings of target legislator and legislation, and concatenate them to encode the relation. Then, we feed the relation representation into a feed-forward neural network (FFNN) with softmax to predict the result. Cross entropy loss is used:

$$\mathcal{L}_{vote} = -\sum_{m,l,k} y_{m,l,k} \log(f_k(m,l)) \qquad (4)$$

where $y_{m,l,k}$ is the k_{th} one-hot class label of legislator m's vote on legislation l and f_k indicates the k_{th} component of the output of activation layer $\sigma(\cdot)$.

3.3.2 Hashtag Usage Prediction

Similar to roll call vote prediction, hashtag usage prediction is modeled as a relation prediction task. The representation of an edge is produced by embeddings of target legislator and hashtag. We then feed this representation to another FFNN with softmax. Cross entropy loss is used:

$$\mathcal{L}_{hashtag} = -\sum_{m,t,k} y_{m,t,k} \log(g_k(m,t)) \qquad (5)$$

where $y_{m,t,k}$ is the k_{th} one-hot post label of legislator m's for hashtag t and g_k indicates the k_{th} component of the output of activation layer $\sigma(\cdot)$.

3.3.3 Following Proximity Loss

Previous studies (Barberá, 2015; Peng et al., 2016) have proved the effectiveness of using the following relationships on Twitter for political preference estimation, and show that users prefer to follow those with similar political positions. In order to incorporate this factor into consideration, we introduce a proximity loss (Hamilton et al., 2017; Nguyen et al., 2020) computed from a following network of legislators. It enables neighboring nodes to be represented more similarly and alienates representations of un-associated nodes. The

proximity loss is formulated as follows:

$$\mathcal{L}_{prox} = - \sum_{m \in G'} \Big(\log \Big(\sigma \Big(e_m^\top e_{m_p} \Big) \Big)$$
$$+ Q \cdot \mathbb{E}_{m_n \sim P_n(m)} \log \Big(\sigma \Big(-e_m^\top e_{m_n} \Big) \Big) \Big) \quad (6)$$

where G' is the subgraph of legislators formed by following relationships, and e_m is the representation of a legislator m. m_p is a neighbor of m that can be derived using fixed-length random walk, while m_n is a negative sample that can be obtained through negative sampling $m_n \sim P_n(m)$ (Hamilton et al., 2017). Q controls the number of negative samples.

We form the final loss by linearly combining these three factors: $\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{vote} + \lambda_2 \mathcal{L}_{hashtag} + \lambda_3 \mathcal{L}_{prox}$, where λ_1, λ_2 and λ_3 are hyperparameters controlling the weight of different losses.

4 Experiments

4.1 Experiment Setup

Dataset Splits Our experiment is based on data from the 112th to 115th congress, including both bills and resolutions from House and Senate. We use two configurations to form the experimental dataset. *(1) random*: We set up an in-session experiment environment following Kornilova et al. (2018); Davoodi et al. (2020), where records of each two-year session is considered as an independent experiment set. This results in 4 experiment sets. For each set, 20% legislation is selected for testing, 20% is for validation and the rest is for training. *(2) time-based*: We set up a time-based environment following Yang et al. (2020). We form an experiment set with two consecutive sessions and use the former one for training and validation and the latter one for testing respectively. This results in 3 experiment sets. In this setting, some legislators might appear in the testing session only. Therefore, we report results of two settings. For *Mem Train*, we only include legislators appearing in training set for testing. For *Mem All*, we include all legislators in test set.

Implementation Details The dimensions of initial legislative representations are 64, 768 and 768 for legislator, legislation and hashtag respectively. We randomly choose 50 tweets to encode each hashtag. When modeling relations, we set a threshold as the mean value for each type of relations, and only

reserve those with weights greater than the threshold, to eliminate noise. We use 2-layer RGCNs and the sizes of hidden layers are 128 and 64. A batch normalization layer is added after initializing representation. The batch size is 128 and learning rate is 1×10^{-4}. Dropout and early stopping strategies are adopted to prevent the model from over-fitting. For hyperparameters of three losses, we simply set $\lambda_1 = \lambda_2 = 10\lambda_3$ to control three losses within the same order of magnitude. For graph construction, the entity set covers all entities involved in and before that year while the relation set only covers information before that year to avoid future information leakage.

Models for Comparison We compare our model with some state-of-the-art approaches.

- *majority* is a baseline which assumes all legislators vote *yea*.

- *ideal-point-wf* (Gerrish and Blei, 2011): a regression model that takes the word frequency of legislation text as features. The training paradigm follows the traditional ideal point model. Thus, it can only predict on legislators present in the training data.

- *ideal-point-tfidf*: similar to *ideal-point-wf*, it uses TFIDF of legislation text as features instead.

- *ideal-vector* (Kraft et al., 2016): it learns multi-dimensional ideal vectors for legislators based on bill texts.

- *CNN* (Kornilova et al., 2018): it uses CNN to encode legislation.

- *CNN+meta* (Kornilova et al., 2018): on the basis of *CNN*, it adds percentage of sponsors of different parties as bill's authorship information.

- *LSTM+GCN* (Yang et al., 2020): it uses LSTM to encode legislation and applies a GCN to update representations of legislators.

- *Vote*: the single task of roll call vote in our framework.

- *Ours*: our framework.

4.2 Overall Performance

We report the average accuracy of all experiment sets following Kornilova et al. (2018); Yang et al. (2020). Besides, macro F1 score is also provided for more information. Table 1 shows the overall performance for roll call vote prediction.

| methods | random | | time-based | | | |
| | | | Mem Train | | Mem All | |
	Acc.	MaF	Acc.	MaF	Acc.	MaF
majority	77.48	43.62	76.16	43.21	77.40	43.62
ideal-point-wf	85.37	78.48	65.72	53.30	-	-
ideal-point-tfidf	86.46	80.02	66.41	54.15	-	-
ideal-vector	87.35	80.15	85.54	79.71	81.95	75.49
CNN	87.28	80.34	85.66	78.90	81.97	75.68
CNN+meta	88.02	81.59	86.40	80.44	84.30	77.67
LSTM+GCN	88.41	82.26	87.01	80.91	85.82	80.73
Vote	90.22	84.92	89.90	84.72	89.76	84.35
Ours	**91.84**	**86.73**	**90.52**	**85.91**	**90.61**	**85.45**

Table 1: Overall performance of different models for roll call vote prediction. *random* stands for in-session setup. *Mem Train* reports performance on legislators appear in training set while *Mem All* reports results on all legislators in test set.

Roll Call Vote Prediction   We have several findings for results of roll call vote prediction.

- Our model yields the best results. By utilizing hashtag usage information, our framework can further improve the performance on the basis of the single task *Vote*.

- Neural networks based approaches perform better than ideal-point based models. *CNN+meta* and *LSTM+GCN* achieve better results than other baselines. This proves that introducing background information is helpful to capture general preferences.

- All models perform worse in *time-based* setting compared to *random* setting. The performance drop of ideal-point based models that incorporate textual information is the largest. This indicates that ideal-point based models have difficulty for transfer learning from one session to another.

- Comparing the setting of *Mem Train* and *Mem All*, we find that most methods have difficulty modeling new-elected legislators. Models incorporating background knowledge perform more stable, among which our model is the most robust one.

Hashtag Usage Prediction   For hashtag usage prediction, we evaluate our model in *time-based* setting. For comparison, we employ a simple FFNN to process initial embeddings of legislators and hashtags for label prediction. Experiment results show that our model achieves better performance than FFNN in terms of both accuracy (80.44% vs 80.03%) and macro F1 (61.34% vs 53.93%). This indicates that it's difficult to predict preferences on hashtags of legislators based on textual information only. Incorporating legislative information, our model achieves improvements, especially for *macro F1*. This also demonstrates that learning the voting behavior of legislators also benefits predicting what they will say.

4.3   Influence of Noise in Hashtag Set

Although most hashtags are polarized, there are still general ones like *#America* and *#Trump* . The usage of these hashtags is not able to stand for the stance. Therefore, the set of hashtags in our dataset contains noise. We conduct an additional experiment to explore the influence of noise brought by hashtags on the task of roll call prediction. We set a threshold to filter noise. Different thresholds indicate different degrees of polarization, where 0.5 means using all hashtag labels in our dataset (the setting of our model in Table 1), and 0.8 represents the ratio of major sentiment in tweets of the hashtag must exceed 0.8. Figure 4a presents the results. The performance increases when the threshold increases from 0.5 to 0.7, indicating hashtags without firm attitudes would hurt the performance. After that, the performance drops because of the reduction of data. However, due to the chance of hashtag hijacking strategy where a hashtag is deliberately taken up and used by "the other side"(Hadgu et al., 2013), noise in hashtags can not be completely eliminated in this way.
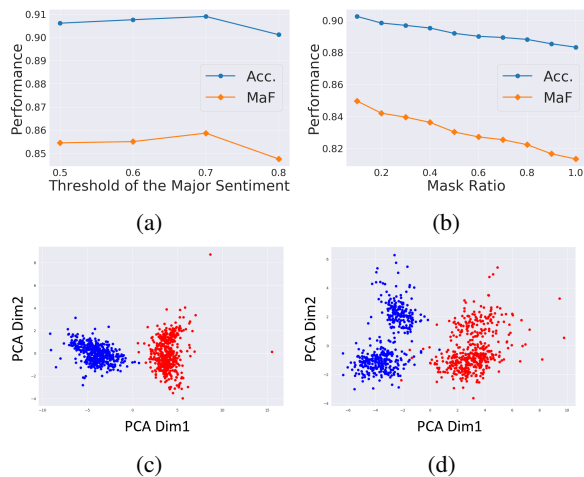
Figure 4: Further analysis on the experiment results. (a) influence of hashtags used on the performance. (b) cold start simulation. (c) visualization of legislator representation without hashtag prediction. (d) legislator representation of our model.

5 Further Analysis

We perform additional analysis to further evaluate the effectiveness of our model.

5.1 Cold Start Simulation

Since our model makes use of statements on Twitter to densify connections among legislators, we want to explore its ability to deal with the cold start problem. Although the settings of *Mem Train* and *Mem All* have shown the advantage of our model for newly-elected legislator modeling, we set up a more general environment. Here, we randomly mask a certain ratio of legislators, that is, discard their historical legislative information when constructing graph, to better investigate the model's ability to mitigate the cold start problem. Figure 4b illustrates the performance of our model when masking different ratios of legislators in time-based setting. When the ratio increases, performance stays stable and performs better than the best baseline *LSTM+GCN* consistently (87.01% of Acc. and 80.91% of MaF.). Thus, taking advantage of content generated by legislators, our proposed model shows good robustness.

5.2 Legislator Representation

We project learned representation of legislators into a 2D space using PCA. Figure 4c shows legislator representation of 115th congress based on data of 2018 learned by vote-based model, i.e., to train our framework without hashtag information. Figure
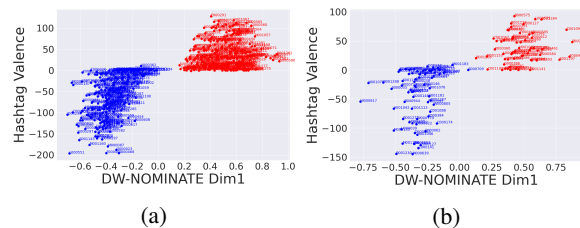

Figure 5: Comparing hashtag valence and DW-NOMINATE Dim1. (a) House. (b) Senate.

4d shows that learned by the overall framework, where Democrats clearly fall into two clusters. An explanation can be given with a closer look at the relations between legislators and hashtags. While the left lower group behaves actively on Twitter, posting hashtags like *#trumpcare*, *#goptaxscam* and *#protectourcare* for multiple times, the other group rarely expresses their position by using these hashtags. While they vote similarly, this divergence can not be captured relying only on votes. Thus, our method indeed learns nuances between legislators.

5.3 Consistency of Statement and Behavior

We follow Hemphill et al. (2013) to investigate legislators' overall tweeting behavior and voting behavior by comparing hashtag usage and the first dimension of DW-NOMINATE (Lewis and Poole, 2004). We compute hashtag valence proposed by Conover et al. (2011a) and aggregate hashtags a legislator has posted to get hashtag valence for him or her. Since DW-NOMINATE scores are not comparable across chambers, Figure 5a and Figure 5b show conditions for legislators involved in the 115th session of House and Senate respectively. The figures and correlation($r(529) = 0.80$ $p < 0.001$ for House and $r(135) = 0.74$ $p < 0.001$ for Senate) not only indicate that most legislators are polarized similarly in tweeting and voting, but also again illustrate that some legislators voting similarly on average can be hugely different in their languages. Complex similarities and differences of legislators like this can not be expressed by representation learned from votes or tweets separately.

Besides overall leaning inference, inconsistency at the level of individual bills is also worthy of attention. When predicting on *113S2223*, a bill for "an increase in the Federal minimum wage", the vote-based model predicts that Senator Harry Reid will vote *nay*, which is also the ground truth. But our model wrongly predicts that he will vote *yea*. We probe into his tweets and find that he

used *#raisethewage* frequently to call for raise in minimum wage, as those who support the bill. On the one hand, hashtags may have difficulty capturing more fine-grained decisions, which can be influenced by various factors; on the other hand, legislators may behave differently from what they say, since they may make certain statements to get public support (Spell et al., 2020). When legislators do not accord their words to deed, our model may be misled by legislators' statements. As it's difficult to find hashtags directly and accurately related to a specific bill in an automatic and complete way, we will explore the frequency of inconsistency in the future.

6 Related Work

Ideal point estimation has become a mainstream approach to model ideology of legislators. Classical ideal point model (Clinton et al., 2004) represents both legislators and legislation in the same space, and voting behavior is characterized as the distance between them. However, this simple spatial model fails to predict votes on new legislation. Text-based models have emerged to address this issue. Gerrish and Blei (2011, 2012); Gu et al. (2014); Nguyen et al. (2015) extended ideal point model with latent topics and issue-adjusted methods. Some embedding methods (Kraft et al., 2016) also promote learning of legislators. More recently, external context information including party, sponsor and donors (Kornilova et al., 2018; Yang et al., 2020; Davoodi et al., 2020) have been introduced to better describe the legislative process.

Since votes are not the only way to express political preferences, other sources of data including speech and knowledge graph (Budhwar et al., 2018; Gentzkow et al., 2019; Patil et al., 2019; Vafa et al., 2020) have been applied to estimate ideology. Although previous studies (Bruns and Highfield, 2013; Golbeck and Hansen, 2014; Barberá, 2015; Peng et al., 2016; Wong et al., 2016; Boutyline and Willer, 2017; Johnson et al., 2017) have incorporated social network of following or retweeting on Twitter to learn legislators, fine-grained attitudes of legislators remain unknown since the texts themselves have not been mined. Until recently, Preoţiuc-Pietro et al. (2017) started to analyze linguistic differences between ideologically different groups using a broad range of handcrafted language features, and studies (Vafa et al., 2020; Spell et al., 2020) explored to incorporate Twitter texts to cap-

ture nuances in legislators' preferences via statistical methods. In spite of this, there has been little research attempting to combine votes with public statements to portray legislators from both angles and predict their behavior.

Previous studies (Conover et al., 2011b; Small, 2011; Bruns and Stieglitz, 2012; Cohen and Ruths, 2013) have suggested that modeling on hashtag metadata is an informative way to analyze tweets, yielding classification of political affiliations. Since hashtag is an important mean for people to participate in political discussion and communication, hashtag usage pattern has also been modeled as feature vectors in many clustering tasks to help learn different user groups (Conover et al., 2011a; Bode et al., 2013, 2015). Hemphill et al. (2013) and Yang et al. (2016) have analyzed hashtag usage patterns of different ideologies through feature selection and keyword statistics. However, hashtag usage can be further utilized based on these analyses, e.g., for prediction tasks. Thus, we focus on hashtags to depict statements of legislators on Twitter, to jointly estimate their political preferences.

7 Conclusions and Future Work

In this paper, we take the first step to align voting behavior with statements on Twitter to jointly learn representation of legislators. We construct a heterogeneous graph to model the legislative context with a hashtag usage prediction task proposed to jointly train. Experiments demonstrate that our framework can learn effective legislative representation and yield improvements for the roll call vote prediction task. Due to the deficiency of background information, we have not yet detected more fine-grained stance of legislators towards specific events. In the future, we aim to conduct more research on the stance modeling of legislators.

Acknowledgments

References

Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, 23(1):76–91.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482.

Leticia Bode, Alexander Hanna, Ben Sayre, JungHwan Yang, and Dhavan V Shah. 2013. Mapping the political twitterverse: Finding connections between political elites.

Leticia Bode, Alexander Hanna, Junghwan Yang, and Dhavan V Shah. 2015. Candidate networks, citizen clusters, and political expression: Strategic hashtag use in the 2010 midterms. *The ANNALS of the American Academy of Political and Social Science*, 659(1):149–165.

Andrei Boutyline and Robb Willer. 2017. The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political psychology*, 38(3):551–569.

Axel Bruns and Tim Highfield. 2013. Political networks on twitter: Tweeting the queensland state election. *Information, Communication & Society*, 16(5):667–691.

Axel Bruns and Stefan Stieglitz. 2012. Quantitative approaches to comparing communication patterns on twitter. *Journal of technology in human services*, 30(3-4):160–185.

Aditya Budhwar, Toshihiro Kuboi, Alex Dekhtyar, and Foaad Khosmood. 2018. Predicting the vote using legislative speech. In *Proceedings of the 19th annual international conference on digital government research: governance in the data age*, pages 1–10.

Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.*, 10:103–126.

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

Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on twitter: It's not easy! In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7.

Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011a. Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5.

Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011b. Predicting the political alignment of twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 192–199. IEEE.

Maryam Davoodi, Eric Waltenburg, and Dan Goldwasser. 2020. Understanding the language of political agreement and disagreement in legislative texts. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5358–5368, Online. Association for Computational Linguistics.

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 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

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

Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Sean Gerrish and David Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems*, volume 25, pages 2753–2761. Curran Associates, Inc.

Sean M Gerrish and David M Blei. 2011. Predicting legislative roll calls from text. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*.

Jennifer Golbeck and Derek Hansen. 2014. A method for computing political preference among twitter followers. *Social Networks*, 36:177–184.

Yupeng Gu, Yizhou Sun, Ning Jiang, Bingyu Wang, and Ting Chen. 2014. Topic-factorized ideal point estimation model for legislative voting network. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 183–192.

Asmelash Teka Hadgu, Kiran Garimella, and Ingmar Weber. 2013. Political hashtag hijacking in the us. In *Proceedings of the 22nd international conference on World Wide Web*, pages 55–56.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034.

Libby Hemphill, Aron Culotta, and Matthew Heston. 2013. Framing in social media: How the us congress uses twitter hashtags to frame political issues. *SSRN Electronic Journal*.

Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 741–752.

Hamid Karimi, Tyler Derr, Aaron Brookhouse, and Jiliang Tang. 2019. Multi-factor congressional vote prediction. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 266–273.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Anastassia Kornilova, Daniel Argyle, and Vladimir Eidelman. 2018. Party matters: Enhancing legislative embeddings with author attributes for vote prediction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 510–515, Melbourne, Australia. Association for Computational Linguistics.

Peter Kraft, Hirsh Jain, and Alexander M. Rush. 2016. An embedding model for predicting roll-call votes. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2066–2070, Austin, Texas. Association for Computational Linguistics.

Jeffrey B. Lewis, Poole Keith, Rosenthal Howard, Boche Adam, Rudkin Aaron, and Sonnet Luke. 2021. Voteview: Congressional roll-call votes database.

Jeffrey B Lewis and Keith T Poole. 2004. Measuring bias and uncertainty in ideal point estimates via the parametric bootstrap. *Political Analysis*, pages 105–127.

Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1165–1174.

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 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448.

Pallavi Patil, Kriti Myer, Ronak Zala, Arpit Singh, Sheshera Mysore, Andrew McCallum, Adrian Benton, and Amanda Stent. 2019. Roll call vote prediction with knowledge augmented models. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 574–581.

Tai-Quan Peng, Mengchen Liu, Yingcai Wu, and Shixia Liu. 2016. Follower-followee network, communication networks, and vote agreement of the us members of congress. *Communication research*, 43(7):996–1024.

Daniel Preoţiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar. 2017. Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 729–740.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Tamara A Small. 2011. What the hashtag? a content analysis of canadian politics on twitter. *Information, communication & society*, 14(6):872–895.

Gregory Spell, Brian Guay, Sunshine Hillygus, and Lawrence Carin. 2020. An Embedding Model for Estimating Legislative Preferences from the Frequency and Sentiment of Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 627–641, Online. Association for Computational Linguistics.

Keyon Vafa, Suresh Naidu, and David Blei. 2020. Text-based ideal points. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5345–5357, Online. Association for Computational Linguistics.

Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. 2016. Quantifying political leaning from tweets, retweets, and retweeters. *IEEE transactions on knowledge and data engineering*, 28(8):2158–2172.

Xinxin Yang, Bo-Chiuan Chen, Mrinmoy Maity, and Emilio Ferrara. 2016. Social politics: Agenda setting and political communication on social media. In *International conference on social informatics*, pages 330–344. Springer.

Yuqiao Yang, Xiaoqiang Lin, Geng Lin, Zengfeng Huang, Changjian Jiang, and Zhongyu Wei. 2020. Joint representation learning of legislator and legislation for roll call prediction. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 1424–1430. International Joint Conferences on Artificial Intelligence Organization. Main track.