# Understanding Politics via Contextualized Discourse Processing

**Rajkumar Pujari**
Purdue University
rpujari@purdue.edu

**Dan Goldwasser**
Purdue University
dgoldwas@purdue.edu

## Abstract

Politicians often have underlying agendas when reacting to events. Arguments in contexts of various events reflect a fairly consistent set of agendas for a given entity. In spite of recent advances in Pretrained Language Models (PLMs), those text representations are not designed to capture such nuanced patterns. In this paper, we propose a Compositional Reader model consisting of encoder and composer modules, that attempts to capture and leverage such information to generate more effective representations for entities, issues, and events. These representations are contextualized by tweets, press releases, issues, news articles, and participating entities. Our model can process several documents at once and generate composed representations for multiple entities over several issues or events. Via qualitative and quantitative empirical analysis, we show that these representations are meaningful and effective.

## 1 Introduction

Often in political discourse, the same argument trajectories are repeated across events by politicians and political caucuses. Knowing and understanding the trajectories that are regularly used, is pivotal in contextualizing the comments made by them when a new event occurs. Furthermore, it helps us in understanding their perspectives and predict their likely reactions to new events and participating entities.

In political text, bias towards a political perspective is often subtle rather than explicitly stated (Fan et al., 2019). Choices of mentioning or omitting certain entities or certain attributes can reveal the author's agenda. For example, when a politician tweets "*mass shootings are due to a huge mental health problem*" in reaction to a new shooting event, it is likely that they oppose gun control and support free gun rights, despite not mentioning their stance explicitly.

Our main insight in this paper is that effectively detecting such bias from text requires modeling the broader context of the document. This can include understanding relevant facts related to the event addressed in the text, the ideological leanings and perspectives expressed by the author in the past, and the sentiment/attitude of the author towards the entities referenced in the text. We suggest that this holistic view can be obtained by combining information from multiple sources, which can be of varying types, such as news articles, social media posts, quotes from press releases and historical beliefs expressed by politicians.

Despite recent advances in Pretrained Language Models (PLMs) in NLP (Devlin et al., 2019; Yang et al., 2019; Liu et al., 2019), which have greatly improved word representations via contextualized embeddings and powerful transformer units, such representations alone are not enough to capture nuanced biases in political discourse. Two of the key reasons are: (i) they do not directly focus on entity/issue-centric data and (ii) they contextualize only based on surrounding text but not on relevant issue/event knowledge.

A computational setting for this approach, *combining context and context analysis*, requires two necessary attributes: (i) an input representation that combines all the different types of information meaningfully and (ii) an ability to process all the information together in one-shot.

We address the first challenge by introducing a graph structure that ties together first-person informal (tweets) and formal discourse (press releases and perspectives), third-person current (news) and consolidated (Wikipedia) discourse. These documents are connected via their authors, the issues/events they discuss and the entities that are mentioned in them. As a clarifying example consider the partial Tweet by President Trump *"The NRA is under siege by Cuomo"*. This tweet will be represented in our graph by connecting the text node to the author node (President Trump) and the referenced entity node (New York Gov. Cuomo). These settings are shown in Fig. 1

Then, we propose a novel neural architecture that can process all the information in the graph together in one-shot. The architecture generates a distributed representation for each item in the graph that is contextualized by the representations of others. In our example, this results in a modified representation for the tweet and the entities thus helping us characterize the opinion of President Trump about Governor Cuomo in context of NRA or *guns* in general. Our architecture builds upon the text representations obtained from BERT (Devlin et al., 2019). It consists of an Encoder which combines all the documents related to a given node to generate an initial node representation and a Composer which is a Graph Attention Network (GAT) that composes over the graph structure to generate contextualized node embeddings.

We design two self-supervised learning tasks to train the model and capture structural dependencies over the rich discourse representation, namely predicting *Authorship* and *Referenced Entity* links over the graph structure. The intuition behind the tasks is that the model is required to understand subtle language usage to solve them. *Authorship* prediction requires the model to differentiate between: (i) the language of one author from another and (ii) the language of the author in context of one issue vs another issue. *Referenced Entity* prediction requires the model to understand the language used by an author when discussing a particular entity given the author's historical discourse.

We evaluate the resulting discourse representation via several empirical tasks identifying political perspectives at both article and author levels. Our evaluation is designed to demonstrate the importance of each component of our model and usefulness of the learning tasks. The *Grade Paraphrase* task evaluates our model's ability to consolidate multiple documents, of different types, from a single author into a coherent perspective about an issue. This is evaluated by framing the problem as a paraphrasing task, comparing the model's composed representation of an author with a short text expressing the stance directly, i.e., only based on the model's pre-training process. The *Grade Prediction* and *Bias Prediction* tasks show that our representations capture meaningful information that make them highly effective for political prediction tasks. Both tasks build classifiers on top of the model. *Grade Prediction* evaluates author and issue representations while *Bias Prediction* evaluates graph-contextualized document representations. We perform *Grade Prediction* for two domains: *guns* and *environment* using politician grades from two different organizations: National Rifles Association (NRA) and League of Conservation Voters (LCV). We compare our model to three competitive baselines: BERT (Devlin et al., 2019), an adaptation of BERT to our data, and our Encoder architecture. This helps us evaluate different aspects of our model as well as our learning tasks. We also analyse the relative usefulness of various types of documents via an ablation study. The BERT adaptation baseline is designed to be trained on our learning tasks without using the Composer architecture. It helps demonstrate the effectiveness of our learning tasks and the importance of the Composer architecture. Our model outperforms the baselines on all three evaluation tasks. Finally, we perform qualitative analysis, visualizing entities' stances, demonstrating that our representations effectively capture nuanced political information. To summarise, our research contributions include:

1. Proposing a novel rich graphical structure to connect various types of documents, entities, issues and events.
2. Proposing an effective neural architecture named Compositional Reader to process all the documents in one-shot. Designing two effective learning tasks to train it.
3. Designing & performing quantitative and qualitative evaluation to show that the graph structure, neural architecture and the representations are meaningful and effective.[1]

## 2 Related Work

Due to recent advances in text representations catalysed by Peters et al. (2018), Vaswani et al. (2017) and followed by Devlin et al. (2019), Liu et al. (2019) and Yang et al. (2019), we are now able to create very rich textual representations that are effective in many nuanced NLP tasks. Al-

[1] Our code and data are available at: https://github.com/pujari-rajkumar/compositional_learner

though semantic contextual information is captured by these models, they are not explicitly designed to capture entity/event-centric information. Hence, to solve tasks that demand better understanding of such information (Chen et al., 2019; Biessmann, 2016; Johnson and Goldwasser, 2018; Kornilova et al., 2018), there is a need to create more focused representations.

Of late, several works attempted to solve such tasks (Iyyer et al., 2016; Han et al., 2019; Demszky et al., 2019; Kulkarni et al., 2018; Preoţiuc-Pietro et al., 2017; Diermeier et al., 2012). But, the representations used are usually limited in scope to specific tasks and not rich enough to capture information that is useful across several tasks.

Compositional Reader model, that builds upon Devlin et al. (2019) embeddings and consists of a transformer-based Graph Attention Network inspired from Veličković et al. (2017) and Müller et al. (2019) aims to address those limitations via a generic entity-issue-event-document graph, which is used to learn highly effective representations.

# 3 Data

| Data Type | Count |
|---|---|
| News Events | 367 |
| Authoring Entities | 455 |
| Referenced Entities | $10,506$ |
| Wikipedia Articles | 455 |
| Tweets | $86,409$ |
| Press Releases | $62,257$ |
| Perspectives | $30,446$ |
| News Articles | $8,244$ |
| Total # documents | $187,811$ |
| Average sents per doc | 14.18 |

Table 1: Summary statistics of data

We collected US political text data related to 8 broad topics: *guns, LGBTQ rights, abortion, immigration, economic policy, taxes, middle east & environment*. Data used for this paper was focused on 455 US senators and congressmen. We collected political text data relevant to above topics from 5 sources: press statements by political entities from ProPublica Congress API[2], Wikipedia articles describing political entities, tweets by political entities (Congress Tweets, Baumgartner (2019)), perspectives of the senators and congressmen regarding various political issues from ontheissues.org and news articles & background of the those political issues from allsides.com. A total of $187,811$ documents were used to train our model. Summary statistics are shown in Tab. 1

## 3.1 Event Identification

Event based categorization of documents is performed as follows: news articles related to each issue are ordered by their date of publication. We find the mean ($\mu$) and standard deviation ($\sigma$) of the number of articles published per day for each issue. If more than $\mu + \sigma$ number of articles are published on a single day for a given issue, we flag it as the beginning of an event. Then, we skip 7 days and look for a new event. Until a new event window begins, the current event window continues. We use thus obtained event windows to mark events.

In our setting, events with in a given issue are non overlapping. We divide events for each issue separately, hence events for different issues overlap. These events last for $7 - 10$ days on average and hence the non-overlapping assumption within an issue is a reasonable relaxation of reality. To illustrate our point: coronavirus and civil-rights are separate issues and hence have overlapping events. An example event related to coronavirus could be "First case of COVID-19 outside of China reporte". Similarly an event about civil-rights could be that "Officer who was part of George Floyd killing suspended". We inspected the events manually and found that the events are meaningful for a high percentage of inspected cases ($\geq 85\%$ events). Examples of identified events are shown in the appendix.

## 3.2 Data Pre-processing

We use Stanford CoreNLP tool (Manning et al., 2014), Wikifier (Brank et al., 2017) and BERT-base-uncased implementation by Wolf et al. (2019) to preprocess data for our experiments. We tokenize the documents, apply coreference resolution and extract referenced entities from each document. The referenced entities are then wikified using Wikifier tool (Brank et al., 2017). The documents are then categorized by issues and events. News articles from allsides.com and perspectives from ontheissues.org are already classified by issues. We use keyword based querying to extract issue-wise press releases from Propublica API. We use hashtag based classification for tweets. A set of gold hashtags for each issue was created and

[2] https://projects.propublica.org/api-docs/congress-api/

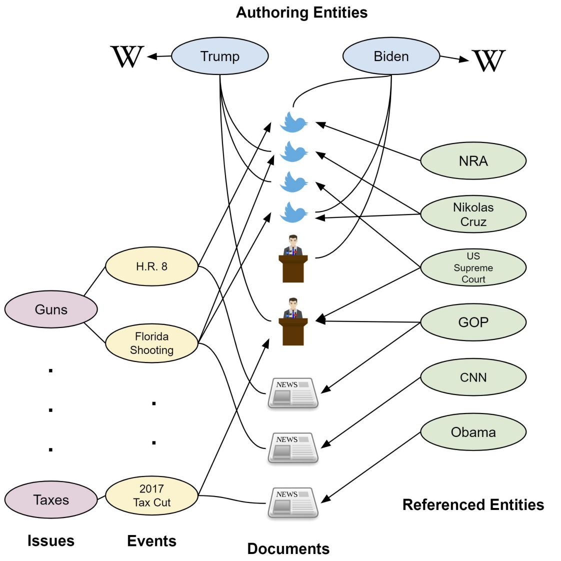


Figure 1: Example Text Graph from Graph Generator

the tweets were classified accordingly[3]. Sentence-wise BERT-base embeddings of all documents are computed.

### 3.3 Query Mechanism

We implemented a query mechanism to obtain relevant subsets of data from the corpus. Each query is a triplet of *entities, issues & lists of event indices corresponding to each of the issues*. Given a query triplet, news articles related to the events for each of the issues, Wikipedia articles for each of the entities, background descriptions of the issues, perspectives of each entity regarding each of the issues and tweets & press releases by each of the entities related to the events in the query are retrieved. Referenced entities for each of the sentences in documents and sentence-wise BERT embeddings of the documents are also retrieved.

## 4 Compositional Reader

In this section, we describe the architecture of the proposed 'Compositional Reader' model in detail. It contains 3 key components: Graph Generator, Encoder and Composer. Given a query output of the query mechanism from Sec. 3.3, Graph Generator creates a directed graph with entities, issues, events and documents as nodes. Encoder is used to generate initial node embeddings for each of the nodes. Composer is a transformer-based Graph Attention Network (GAT) followed by a pooling layer. It generates the final node embeddings and a single summary embedding for the query graph. Each component is described below.

[3]Data collection is detailed in appendix

### 4.1 Graph Generator

Given the output of the query mechanism for a query, the Graph Generator creates a directed graph with 5 types of nodes: authoring entities, referenced entities, issues, events and documents. Directed edges are used by Composer to update source nodes' representations using destination nodes. We design the topology with the main goal of capturing the representations of events, issues and referenced entities that reflect author's opinion about them. We add edges from issues/events to author's documents but omit the other direction as our main goal is to contextualize issues/events using author's opinions.

Edges are added from authoring entities to their Wikipedia articles and the documents authored by it (tweets, press releases and perspectives). Reverse edges from the authored documents to the author are also added. Uni-directional edges from relevant event nodes to the tweet and press release document nodes are added. Edges from issue nodes to event nodes and vice-versa are added. Edges from the issue nodes to their background description documents are added. Edges from event nodes to news articles describing the events and vice-versa are added. Uni-directional nodes from issue nodes to author perspective nodes are added. Finally, uni-directional edges from referenced entities to all the document nodes are added. An example graph is shown in Fig. 1.

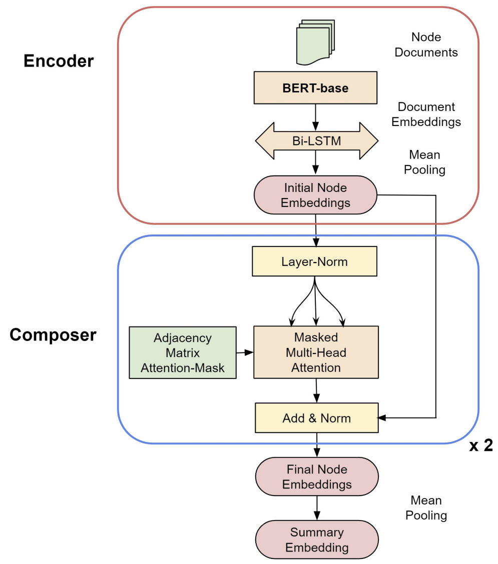


Figure 2: Encoder-Composer Architecture

### 4.2 Encoder

Encoder is used to compute the initial node embeddings. It consists of BERT followed by a Bi-LSTM. For each node, it takes a sequence of documents as input. The documents are ordered temporally. The output of Encoder is a single embedding of dimension $d_m$ for each node. Given a node $\mathcal{N} = \{D_1, D_2, \ldots, D_d\}$ consisting of $d$ documents, for each document $D_i$, contextualized embeddings of all the tokens are computed using BERT. Token embeddings are computed sentence-wise to avoid truncating long documents. Then, token embeddings of each document are mean-pooled to get the document embeddings $\vec{\mathcal{N}}^{bert} = \{\vec{D}_1^{bert}, \vec{D}_2^{bert}, \ldots, \vec{D}_d^{bert}\}$ where $\vec{D}_i^{bert} \in \mathbb{R}^{1\times d_m}$, $d_m$ is the dimension of a BERT token embedding. The sequence $\vec{\mathcal{N}}^{bert}$ is passed through a Bi-LSTM to obtain an output sequence $\vec{E} = \{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_d}\}$, $\vec{e_i} \in \mathbb{R}^{1\times h}$, where $h/2$ is the hidden dimension of the Bi-LSTM, we set $h = d_m$ in our model. Finally, the output of Encoder is computed by mean-pooling the sequence $\vec{E}$. We use BERT-base-uncased model in our experiments where $d_m = h = 768$.

Initial node embeddings of all the document nodes are set to Encoder output of the documents themselves. For authoring entity nodes, their Wikipedia descriptions, tweets, press releases and perspective documents are passed through Encoder. For issue nodes, background description of the issue is used. For event nodes, Encoder representation of all the news articles related to the event is used. For referenced entities, all documents referring to the entity are used.

### 4.3 Composer

Composer is a transformer-based graph attention network (GAT) followed by a pooling layer. We use the transformer encoding layer proposed by Vaswani et al. (2017) after removing the position-wise feed forward layer as a graph attention layer. Position-wise feed forward layer is removed because the transformer unit was originally proposed for sequence to sequence prediction tasks, but the nodes in a graph usually have no ordering relationship between them. Adjacency matrix of the graph is used as the attention mask. Self-loops are added for all nodes so that updated representation of the node also depends on its previous representation. Composer module uses $l = 2$ graph attention layers in our experiments. Composer module generates updated node embeddings $\mathbb{U} \in \mathbb{R}^{n\times d_m}$ and a summary embedding $\mathbb{S} \in \mathbb{R}^{1\times d_m}$ as outputs. The output dimension of node embeddings is 768, same as BERT-base.

$$
\begin{aligned}
&\mathbb{E} \in \mathbb{R}^{d_m\times n}, \mathcal{A} \in \{0,1\}^{n\times n}\\
&\mathbb{G} = layer-norm(\mathbb{E})\\
&Q = W_q^T \mathbb{G}, Q \in \mathbb{R}^{n_h\times d_k\times n}\\
&K = W_k^T \mathbb{G}, K \in \mathbb{R}^{n_h\times d_k\times n}\\
&V = W_v^T \mathbb{G}, V \in \mathbb{R}^{n_h\times d_v\times n}\\
&M = \frac{Q^T K}{\sqrt{d_k}}, M \in \mathbb{R}^{n_h\times n\times n}\\
&M = mask(M, \mathcal{A})\\
&\mathbb{O} = MV^T, \mathbb{O} \in \mathbb{R}^{n_h d_v\times n}\\
&\mathbb{U} = W_o^T \mathbb{O} + \mathbb{E}\\
&\quad \mathbb{S} = mean-pool(\mathbb{U})
\end{aligned}
\tag{1}
$$

where $n$ is number of nodes in the graph, $d_m$ is the dimension of a BERT token embedding, $d_k$, $d_v$ are projection dimensions, $n_h$ is number of attention heads used and $W_q \in \mathbb{R}^{d_m\times n_h d_k}$, $W_k \in \mathbb{R}^{d_m\times n_h d_k}$, $W_v \in \mathbb{R}^{d_m\times n_h d_v}$ and $W_o \in \mathbb{R}^{n_h d_v\times d_m}$ are weight parameters to be learnt. $\mathbb{E} \in \mathbb{R}^{d_m\times n}$ is the outputs of the encoder. $\mathcal{A} \in \{0,1\}^{n\times n}$ is the adjacency matrix. We set $n_h = 12$, $d_k = d_v = 64$ in our experiments.

## 5 Learning Tasks

We design 2 learning tasks to train the Compositional Reader model: Authorship Prediction and Referenced Entity Prediction. Both the tasks are different flavors of link prediction over graphs. In Authorship Prediction, given a graph, an author node and a document node with no link between them, the task is to predict if the document was authored by the author node. In the Referenced Entity Prediction task, given a graph, a document node and a referenced entity node, the task is to predict if the entity was referenced in the document. For this task, all occurrences of one entity in the text are replaced with a generic <ent> token in the document text before the document embedding is computed. Both are detailed below.

### 5.1 Authorship Prediction

Authorship Prediction is designed as a binary classification task. In this task, given a graph generated by the graph generator model $\mathcal{G}$, an author node $n_a$ and a document node $n_d$ with no edges between them, the task is to predict whether or not author represented by node $n_a$ authored the document represented by node $n_d$.

Intuition behind this learning task is to enable our model to learn differentiating between the language of an author in context of an issue and documents by other entities or documents related to other issues. The model sees documents by the same author for the same issue in the graph and learns to decide whether the input document has similar language or not. It is a fairly simple learning task and hence is an ideal task to start pre-training our model.

**Architecture** We concatenate the initial and final node embeddings of the author, document and also the summary embedding of the graph to obtain inputs to the fine-tuning layers for Authorship Prediction task. We add one hidden layer of dimension $384$ before the classification layer.

**Data** Data samples for the task were created as follows: for each of the $455$ entities, for each of the $8$ issues and for all events related to that issue, we fire a query to the query mechanism and use the graph generator module to obtain a data graph (Fig. 1). Hence, we fire $3,640$ queries in total and obtain respective data graphs. To create a positive data sample, we sample a document $d_i$ authored by the entity $a_i$ and remove the edges between the nodes that represent the $a_i$ and $d_i$. Negative samples were designed carefully in $3$ batches to enable the model to learn different aspects of the language used by the author. In the first batch, we sample news article nodes from the same graph. In the second batch, we obtain tweets, press releases and perspectives of the same author but from a different issue. In the third batch, we sample documents related to the same issue but from other authors.

We generate $421,284$ samples in total, with $252,575$ positive samples and $168,709$ negative samples. We randomly split the data into training set of $272,159$ samples, validation set of $73,410$ samples and test set of $75,715$ samples.

**Out-sample Evaluation** We also perform out-sample experiments to evaluate generalization capability to unseen politicians' data. We train the model on training data from two-thirds of politicians and test on the test sets of others. Results are shown in Tab. 3.

**Graph Trimming** We perform graph trimming to make the computation tractable on a single GPU. We randomly drop $80\%$ of the news articles, tweets and press releases that are not related to the event to which $d_i$ belongs. We use graphs with 200-500 nodes and batch size of $1$.

### 5.2 Referenced Entity Prediction

This task is also designed as binary classification. Given a graph $\mathcal{G}$, document node $d_i$ and referenced entity node $r_i$ from $\mathcal{G}$, the task is to predict whether or not $r_i$ is referenced in $d_i$. To create data samples for this task, we sample a document from the data graph, replace all occurrences of the most frequent referenced entity in the document with a generic $<$`ent`$>$ token. We remove the link between $r_i$ and $d_i$ in $\mathcal{G}$. Triplet ($\mathcal{G}$, $d_i$, $r_i$) is used as a positive data sample. We sample another referenced entity $r_j$ from the graph, that is not referenced in $d_i$, to generate a negative sample.

Intuition behind this learning task is to enable our model to learn the correlation between the author, language in the document and the referenced entity. For example, in context of recent Donald Trump's impeachment hearing, consider the sentence 'X needs to face the consequences of their actions'. Depending upon the author, X could either be '*Donald Trump*' or '*Democrats*'. Learning to understand such correlations by looking at other documents from the same author is a useful training task for our model. This is also a harder learning problem than Authorship Prediction.

**Architecture** We use fine-tuning architecture similar to Authorship Prediction on top of Compositional Reader for this task as well. We keep separate fine-tuning parameters for each task as they are fundamentally different prediction problems. Compositional Reader is shared.

**Data** We generated $252,578$ samples for this task, half of them positive. They were split into $180,578$ training samples, validation and test sets of $36,400$ samples each. We apply graph trimming for this task as well. We also perform out-sample evaluation for this learning task.

## 6 Evaluation

We evaluate our model and pre-training tasks in a systematic manner using several quantitative tasks and qualitative analysis. Quantitative evaluation includes 'NRA Grade Paraphrase' task, 'Grade Prediction' on NRA and LCV grades data followed by 'Bias Predication' task on AllSides news articles. Qualitative evaluation includes entity-stance visualization for issues. We compare our model's performance to BERT representations, the BERT adaptation baseline and representations from the Encoder module. Baselines and the evaluation tasks are detailed below.

### 6.1 Baselines

**BERT:** We compute the results obtained by using pooled BERT representations of relevant documents for each of the quantitative tasks. Details of the chosen documents and the pooling procedure is described in the relevant task subsections.

**Encoder Representations:** We compare the performance of our model to the results obtained by using initial node embeddings generated from the Encoder for each of the quantitative tasks.

**BERT Adaptation Model:** We design a BERT adaptation baseline for the learning tasks. BERT adaptation is equivalent to using only the Encoder's initial node embeddings of the Compositional Reader model. While BERT adaptation and Encoder share exactly the same architecture, Encoder parameters are trained via back-propagation through the Composer, while BERT adaptation parameters are trained directly using our learning tasks. In BERT adaptation, once we generate the data graph, we pass the mean-pooled sentence-wise BERT embeddings of the node documents through a Bi-LSTM. We mean-pool the output of Bi-LSTM to get node embeddings. We use fine-tuning layers on top of thus obtained node embeddings for Authorship Prediction and Referenced Entity Prediction tasks. BERT Adaptation baseline allows us to showcase the importance of our proposed training tasks via comparison with Devlin et al. (2019) representations as well as the effectiveness of our Composer architecture in comparison to Compositional Reader model.

### 6.2 NRA Grades Evaluation

National Rifle Association (NRA) assigns letter grades (A+, A, ..., F) to politicians based on candidate questionnaire and their gun-related voting. We evaluate our representations on their ability to predict these grades. Our intuition behind this evaluation is that the language in the tweets, press releases and perspectives of a politician directly helps in predicting their NRA grade. We evaluate our model on 2 tasks, namely, 'Paraphrase Task' and 'Grades Prediction Task'. In the Paraphrase task, we evaluate the representations from our model directly without training on NRA grades data. In the Grade Prediction task, we use the representations from our model and fine-tune on grades data.

We collected the historical data of politicians' NRA grades from everytown.org. Grade data is available for 349 out of 455 politicians in focus. For each politician $p_i$, we obtain data for the query ($p_i$, *guns*, all *guns*-related events). We input the data to Compositional Reader and take the final node embeddings of nodes representing the politician ($\vec{n}_{auth}$), issue *guns* ($\vec{n}_{guns}$) and referenced entity *NRA* ($\vec{n}_{NRA}$). For some politicians, $\vec{n}_{NRA}$ is not available, depending on whether or not they referred to NRA in their discourse. These embeddings are used for both the prediction and paraphrase tasks. We repeat the 'Grade Prediction' task with grades from 'League of Conservation Voters' data for the issue *environment*. The tasks are detailed below.

**NRA Grades Paraphrase Task** In this task, we evaluate our representations directly *without* training on the NRA grade data. Grades are divided into two classes: higher than, and including, B+ are in the positive class and all grades from C+ to F are classified as negative. We formulate a representative sentence for each class:

- POSITIVE: *I strongly support the NRA*
- NEGATIVE: *I vehemently oppose the NRA*

We compute BERT embeddings for the representative sentences to obtain $\vec{pos}_{NRA}$ and $\vec{neg}_{NRA}$. We mean-pool the three embeddings $\vec{n}_{auth}$, $\vec{n}_{guns}$ and $\vec{n}_{NRA}$ to obtain $\vec{n}_{stance}$. We compute cosine similarity of $\vec{n}_{stance}$ with $\vec{pos}_{NRA}$ & $\vec{neg}_{NRA}$. Politician is assigned the higher similarity class.

We compare our model's results to Devlin et al. (2019), BERT adaptation and Encoder embeddings. For Devlin et al. (2019), we compute $\vec{n}_{stance}$ by mean-pooling the sentence-wise BERT embeddings of tweets, press releases and perspectives of the author on all events related to the issue *guns*. Results are shown in Tab. 2.

**NRA Grade Prediction Task** This is as a 5-class classification task, one class for each letter grade: {A, B, C, D & F}. We train a simple feed-forward network with 1 hidden layer of dimension 1000. The network is given 2 inputs $\vec{n}_{auth}$ & $\vec{n}_{guns}$. When $\vec{n}_{NRA}$ is available for an entity, we set $\vec{n}_{guns}$ = *mean*($\vec{n}_{NRA}$, $\vec{n}_{guns}$). The network's output is a classification prediction.

We randomly divide the NRA Grades data into $k = 10$ folds and we train the model with 8 folds and check the performance on $1$ test fold. We use $1$ fold for validation. We repeat this experiment with each fold as the test fold and then the entire process for $5$ random seeds.

We perform this evaluation for Devlin et al. (2019), BERT adaptation, Encoder and Compositional Reader. To compute $\vec{n}_{auth}$ for (Devlin et al., 2019), we mean-pool the sentence-wise embeddings of all author documents on *guns*. For $\vec{n}_{guns}$, we use the background description document of issue *guns*. Results on the test set are in Tab. 2.

Further, we also perform experiments by training the model on a fraction of the data. We monitor the validation and test performances with change in training data percentage. We observe that, in general, the gap between Compositional Reader model and the BERT baseline widens with increase in training data. It hints that our representation likely captures more relevant information for this task. Results are included in the Appendix.

### 6.3 LCV Grade Prediction Task

This is similar to NRA Grade Prediction task. It is a $4$-way classification task. LCV assigns a score ranging between 0-100 to each politician depending upon their environmental voting activity. We segregate politicians into $4$ classes ($0-25$, $25-50$, $50-75$, $75-100$). We obtain input to the prediction model by concatenating $\vec{n}_(auth)$ and $\vec{n}_{environment}$. We use same fine-tuning architecture as NRA Grade Prediction task with a fresh set of parameters. Results are shown in Tab. 2

### 6.4 Bias Prediction in News Articles

In this task, we evaluate the ability of the graph-contextualized representations of the documents to predict bias in news articles. This tasks evaluates the usefulness of the composer architecture in enriching the representations of the documents by propagating information via the referenced entity nodes. We use news articles collected from AllSides for this task. These articles are different from the ones used in our learning tasks. The news displayed on AllSides is labeled left/right/center leaning by the website. We create an issue node, all news articles related to the issue and all the entities that are referenced in the news articles. We initiate the embeddings of the news articles with mean-pooled sentence-wise BERT embeddings of the articles. We use the description from OnTheIssues for the issue node. Then, we compute updated representations for the articles by running the encoder-composer architecture on the graph. We use the updated representations for 3-way bias prediction task. We don't train the encoder-composer parameters in this task. We use $5,828$ training, $979$ validation and $354$ test examples. Results for this task are show in Tab 2.

### 6.5 Opinion Descriptor Generation

This task demonstrates a simple way to interpret our contextualized representations as natural language descriptors. It is an unsupervised qualitative evaluation task. We generate opinion descriptors for authoring entities for specific issues. We use the final node embedding of the issue node ($\vec{n}_{issue}$) for each politician to generate opinion descriptors. Inspired from Han et al. (2019), we define our candidate space for descriptors as the set of adjectives used by the entity in their tweets, press releases and perspectives related to an issue. Although Han et al. (2019) uses verbs as relationship descriptor candidates, we opine that adjectives describe opinions better. We compute the representative embedding for each descriptor by mean-pooling the contextualized embeddings of that descriptor from all its occurrences in the politician's discourse. This is the one of the key differences with prior descriptor generation works such as Han et al. (2019) and Iyyer et al. (2016). They work in a static word embedding space. But, our embeddings are contextualized and also reside in a higher dimensional space. In an unsupervised setting, this makes it more challenging to translate from distributional space to natural language tokens. Hence, we restrict the candidate descriptor space more than Han et al. (2019) and Iyyer et al. (2016). We rank all the candidate descriptors according to cosine similarity of its representative embedding with the vector $\vec{n}_{issue}$. We present some of the results in Tab. 5. In contrast to Iyyer et al. (2016) and Han et al. (2019), our model doesn't need the presence of both the entities in text to generate opinion descriptors. This is often the case in first person discourse. Results are shown in table 5.

**Comparison to RMN and LARN** Han et al. (2019) and Iyyer et al. (2016) both take a set of documents and entity pairs as inputs and generate relationship descriptors for the entity pairs in an unsupervised setting. They are both trained in an encoder-decoder style training process in an unsupervised manner. Given new text with an en-

| Model | Paraphrase All Grades | Paraphrase A/F Grades | NRA Test Acc | LCV Test Acc | Bias Pred Test Acc | Bias Pred Test F1 |
|---|---|---|---|---|---|---|
| BERT | 41.55% | 38.52% | $54.83 \pm 1.79$ | $52.63 \pm 1.21$ | $48.31 \pm 0.04$ | $31.47 \pm 0.04$ |
| BERT Adap. | 37.54% | 42.62% | $69.95 \pm 3.33$ | $59.09 \pm 1.77$ | $50.11 \pm 0.01$ | $34.25 \pm 0.00$ |
| Encoder | 56.16% | 48.36% | $81.34 \pm 0.86$ | $63.42 \pm 0.35$ | $44.80 \pm 0.05$ | $30.47 \pm 0.04$ |
| Comp. Reader | 63.32% | 63.93% | $81.62 \pm 1.23$ | $62.24 \pm 0.56$ | $56.95 \pm 0.03$ | $41.52 \pm 0.02$ |

Table 2: Results of Quantitative Evaluation Tasks. Acc denotes Accuracy, NRA and LCV denote Grade Prediction tasks. Mean $\pm$ Std. Dev for 5 random seeds for Grade Prediction and 10 random seeds for Bias Prediction task.

tity pair, they generate $d$ descriptor embeddings that are used to rank candidate descriptors. Iyyer et al. (2016) uses entire vocabulary space while Han et al. (2019) uses 500 most frequent verbs.

In contrast, our model doesn't need the presence of both the entities in text to generate opinion descriptors. This often tends to be the case in tweets and press releases as they are generated directly by the author (first-person discourse). Our model is also capable of summarizing over multiple documents and generating descriptors for several referenced entities and issues at once while they deal with one entity-pair at a time.

## 7 Results

In this section, we present the results of the learning tasks, followed by the quantitative and qualitative evaluation results. Results in table 2 show the usefulness of various components of the architecture. BERT adaptation shows the effectiveness of our learning tasks, while Encoder results show that the same architecture when trained along with Composer generates better representations. Compositional Reader results show the effectiveness of our entire model. Further, qualitative evaluation shows that our embeddings capture meaningful information about entities and issues both.

### 7.1 Learning Tasks

First, we present the results of Authorship Prediction and Referenced Entity Prediction tasks in tables 3 & 4 respectively. Compositional Reader outperforms BERT adaptation baseline on all metrics. On Authorship Prediction, out-sample performance doesn't drop for either model, validating our graph formulation which allows the model to learn linguistic nuances as opposed to over-fitting. On Referenced Entity Prediction, F1 score for our model improves from 77.51 from in-sample to 78.62 on out-sample while BERT adaptation baseline's F1 drops slightly from 75.21 to 73.67.

| Model | IS Acc | IS F1 | OS Acc | OS F1 |
|---|---|---|---|---|
| BERT Adap. | 93.01 | 92.31 | 95.56 | 95.20 |
| Comp. Reader | 99.49 | 99.47 | 99.42 | 99.39 |

Table 3: Authorship Prediction In-Sample & Out-Sample Results on Test Data. Acc Denotes Accuracy and F1 Score for Positive Class is Reported.

| Model | IS Acc | IS F1 | OS Acc | OS F1 |
|---|---|---|---|---|
| BERT Adap. | 76.57 | 75.21 | 76.26 | 73.67 |
| Comp. Reader | 78.52 | 77.51 | 78.98 | 78.62 |

Table 4: Referenced Entity Prediction In-Sample & Out-Sample Results on Test Data. Acc Denotes Accuracy and F1 Score for Positive Class is Reported.

### 7.2 Quantitative Evaluation

**Grade Paraphrase** Further, we present the results of NRA Grade Paraphrase Task in Tab. 2. Representations from Compositional Reader achieve 63.32% accuracy. If we use only Encoder output, we get 56.16%. Mean-pooled BERT-base embeddings get 41.55%. Using node embeddings from BERT adaptation model yields 37.54%. When we evaluate using only 'A' or 'F' grades, we obtain 63.93% accuracy for Compositional Reader, 48.36% for Encoder, 42.62% for BERT adaptation and 38.52% for mean-pooled BERT.

**Grade Prediction** Results of Grade Prediction task are shown in Tab. 2. On *NRA Grade Prediction*, which is a 5-way classification task, our model achieves an accuracy of $81.62 \pm 1.23$ on the test set. Our model outperforms BERT representations by $26.79 \pm 3.02$ absolute points on the test set. On *LCV Grade Prediction* task which is a 4-way classification, our model achieves $9.61 \pm 1.77$ point improvement over BERT representations.

**Bias Prediction** The results are shown in Tab. 2. Compositional Reader achieves $8.64 \pm 0.07$ point test accuracy improvement over BERT em-

| Issue | Opinion Descriptors | Issue | Opinion Descriptors |
|---|---|---|---|
| Mitch McConnell | Republican | Nancy Pelosi | Democrat |
| *abortion* | fundamental, hard, eligible, embryonic, unborn | *abortion* | future, recent, scientific, technological, low |
| *environment* | achievable, more, unobjectionable, favorable, federal | *environment* | forest, critical, endangered, large, clear |
| *guns* | substantive, meaningful, outdone, foreign, several | *guns* | constitutional, ironclad, deductible, unlawful, fair |
| *immigration* | federal, sanctuary, imminent, address, comprehensive | *immigration* | immigrant, skilled, modest, overall, enhanced |
| Donald Trump | Republican | Joe Biden | Democrat |
| *guns* | terrorist, public, ineffective, huge, inevitable, dangerous | *guns* | banning, prohibiting, ban, maintaining, sold |
| *immigration* | early, dumb, birthright, legal, difficult | *taxes* | progressive, economic, across-the-board, annual, top |

Table 5: Opinion Descriptor Labels for Politicians. They show the most representative adjectives used by the politicians in context of each issue.

beddings of the documents on this task. The task is a 3-way classification task. The classes are imbalanced with fewer examples for center articles, hence we reported the macro-F1 scores.

### 7.3 Qualitative Evaluation

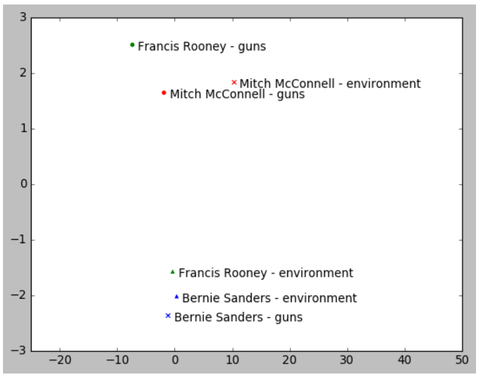


Figure 3: Comparison of Politician Stances on Issues

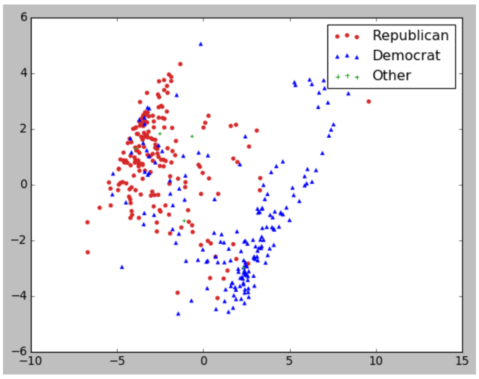


Figure 4: PCA Visualization of issue *guns*

**Politician Visualization** We perform Principle Component Analysis (PCA) on issue embeddings ($\vec{n}_{issue}$) of politicians obtained using the same method as in NRA Grade prediction. We show one such interesting visualization in Fig. 3. Mitch McConnell is a Republican who expressed right-wing views on both *environment* and *guns*. Bernie Sanders is a Democrat that expressed left-wing views on both. Francis Rooney is a Republican who expressed right-wing views on *guns* but left-wing views on *environment*. Fig. 3 demonstrates that this information is captured by our representations. Further examples are in the appendix.

**Issue Visualization** We present visualization of politicians on the issue *guns* in Fig. 4. We observe that *guns* tends to be a polarizing issue. This shows that our representations are able to effectively capture relative stances of politicians. We have included such visualizations for other 7 issues in the appendix. We observe that issues that have traditionally had clear conservative vs liberal boundaries such as *guns* & *LGBTQ rights* are more polarized compared to issues that evolve with time such as *middle-east* & *economic-policy*.

**Opinion Descriptor Generation** We show the results of opinion descriptor generation for few politicians on table 5. These results show the most representative adjectives used by the politicians in context of each of the issues. It can be observed that these descriptors provide a fair reflection of these politicians' views on the issues in focus.

## 8 Ablation Analysis

Further, we investigate the importance of various components of our model. We perform ablation study over various types of documents on the NRA Grades Paraphrase task. the results are shown in Tab. 6 Results in Tab. 6 indicate that *perspectives* are most useful while *tweets* are the least useful documents for the *Grade Paraphrase* task. As *perspectives* are summarized ideological leanings of politicians, it is intuitive that they are more effective for this task. Tweets are informal discourse and tend to be very specific to a current event, hence they are not as useful for this task.

| Model | All Grades |
|---|---|
| Comp.Reader | 63.32% |
| -Tweets | 63.32% |
| -Press Releases | 63.04% |
| -Perspectives | 59.31% |
| Only Tweets | 40.11% |
| Only Press Releases | 55.87% |
| Only Perspectives | 60.74% |

Table 6: Ablation Study on *Grade Paraphrase* task for various types of documents

## 9 Conclusion

We propose a Compositional Reader model that builds upon representations from Devlin et al. (2019) and generates more effective representations. We design learning tasks and train our model on large amounts of political data. We evaluate our model on several qualitative and quantitative tasks. We comprehensively outperform BERT-base model on both learning tasks and quantitative evaluation tasks. Results from our qualitative evaluation demonstrate that our representations effectively capture nuanced political information.

## References


Jason Baumgartner. 2019. Twitter tweets for donald j. trump (@realdonaldtrump).

Felix Biessmann. 2016. Automating political bias prediction. *arXiv preprint arXiv:1608.02195*.

Janez Brank, Gregor Leban, and Marko Grobelnik. 2017. Annotating documents with relevant wikipedia concepts.

Sihao Chen, Daniel Khashabi, Wenpeng Yin, Chris Callison-Burch, and Dan Roth. 2019. Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims. In *Proc. of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

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.

Daniel Diermeier, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. 2012. Language and ideology in congress. *British Journal of Political Science*, 42(1):31–55.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.

Xiaochuang Han, Eunsol Choi, and Chenhao Tan. 2019. No permanent Friends or enemies: Tracking relationships between nations from news. 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 1660–1676, Minneapolis, Minnesota. Association for Computational Linguistics.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former Friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.

Kristen Johnson and Dan Goldwasser. 2018. Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 720–730, Melbourne, Australia. Association for Computational Linguistics.

Anastassia Kornilova, Daniel Argyle, and Vlad Eidelman. 2018. Party matters: Enhancing legislative embeddings with author attributes for vote prediction. *CoRR*, abs/1805.08182.

Vivek Kulkarni, Junting Ye, Steven Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. *CoRR*, abs/1809.03485.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Thomas Müller, Francesco Piccinno, Massimo Nicosia, Peter Shaw, and Yasemin Altun. 2019. Answering conversational questions on structured data without logical forms. *arXiv preprint arXiv:1908.11787*.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

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, Vancouver, Canada. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *CoRR*, abs/1706.03762.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R'emi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5754–5764.

# Appendices

## A Event Examples

In this section, we provide examples of events that were identified by our event identification heuristic. For each automatically extracted event, we observe that the news headlines with in the cluster usually describe the same real world event. The span of each event is 10 days at most. Hence, the assumption that the events with in each issue are non-overlapping is a reasonable relaxation of reality. We made event segregated document data available for future research along with our code.

**Issue - Economic Policy**:

1. **Donald Trump's Tax Proposal Release:**
   - Donald Trump to Propose Tax Breaks on 'Pocketbook' Issues in Economic Plan
   - Trump's economic plan aims to please both corporations and working families
   - Donald Trump Looks to Steady His Campaign With New Economic Speech
   - Trump to outline economic plan in Detroit
   - Clinton to dismiss Trump's economic plan as a 'friends and family discount'
   - OPINION: Trump agenda looks like more of the same
   - Clinton to dismiss Trump's economic plan as a 'friends and family discount'
   - Trump tries to right his campaign, talking of tax cuts
2. **Obama's Economy Speech:**
   - Obama Economy Speech: Why The President's Plan Won't Get Past Republicans
   - U.S. Is 'Through The Worst Of Yesterday's Winds,' Obama Says
   - Obama Blames Five Years of a Bad Economy on "Phony Scandals" and "Distractions"
   - Obama tries to offset current scandals by recycling talking points on economy
   - 5 takeaways from Obama's economy speech
   - Obama at Knox College: 'Washington has taken its eye off the ball
   - Obama: Rest of my presidency is for working-class America
   - Obama Says Private Capital Should Take Lead Mortgage Role
   - Why Obama might tap Summers for Fed despite harsh criticism from left
   - Surprise From Fed: No Pullback In Bond Purchases
   - Obama: Growing income inequality 'defining challenge' of this generation

## B Reproducibility

We use seeds (set to $4056$ for both tasks) for both random example generation and training neural networks. For fine-tuning layers of learning tasks we initialize the models using Xavier uniform (Glorot and Bengio, 2010) initialization with gain=$1.0$. We optimize the parameters using Stochastic Gradient Descent with an initial learning rate=$0.0075$ and momentum=$0.4$. We used $4$ Nvidia GeForce GTX 1080 Ti GPUs with 12 GB memory and linux servers with 64 GB RAM for our experiments. CPU RAM and GPU memory are the main bottlenecks for training the model. It takes $80$ hours to train authorship prediction for $5$ epochs and $14$ hours to train referenced entity prediction task for the same. Generating test results for both tasks together takes $3$ hours. We use a batch size of 1 for both training and evaluation. For NRA Grade Prediction task we use 5 random seeds: $\{5, 7, 11, 13, 17\}$ and report mean and standard deviation. The encoder-composer architecture is made up of $8.26$M parameters, encoder consisting of $3.54M$ and composer $4.72M$. Due to long training time, the only hyperparameter we experimented with is the graph size. We retained as many nodes as possible without exceeding GPU memory ($500$ nodes). Our code and processed data are released[4].

We divide the $3,640$ queries into $151$ batches of $24$ queries each ($3$ politicians $\times$ $8$ issues) and $1$ batch of $16$ queries ($2$ politicians $\times$ $8$ issues). Train, val and test data examples are generated for each query batch. For Authorship Prediction and Referenced Entity Prediction tasks, Compositional Reader model is trained on one batch for $5$ epochs, the best parameters are chosen according to the validation performance of that batch and we proceed to training on future batches. Politicians are ordered randomly when generating queries.

### B.1 Data Collection

We collected data from $5$ sources: Wikipedia, Twitter, ontheissues.org, allsides.com and ProPublica Congress API. We scraped articles from Wikipedia related to all the politicians in focus. We collected tweets from Congress Tweets and Baumgartner (2019). We used a set of hand build gold hashtags to separate them by issues. They are shown at the end of this document. We collected all news articles related to the $8$ issues in focus from allsides.com. We collected press releases from Propublica API using key word search. We use issue names as keywords.

[4]https://github.com/pujari-rajkumar/compositional_learner

## C Gold Hashtags

In this section, we include the gold hashtag set that we built to collect politicians' tweets related to each of the issues.

**Guns**: #endgunviolence, #guncontrol, #gunviolence, #hr8, #nra, #gunsafety, #assaultweaponsban, #gunsense, #marchforourlives, #parkland, #nationalwalkoutday, #disarmhate, #guncontrolnow, #backgroundchecks, #nationalschoolwalkout, #lasvegas, #elpaso, #keepamericanssafe, #gunrights, #erpoact, #lasvegasshooting, #gunreform, #hr1112, #parklandstrong, #elpasostrong, #hr3435, #massshootings, #parklandstudentsspeak

**Taxes**: #GOPTaxScam, #TaxReform, #TaxAndJobsAct, #taxreform, #goptaxscam, #taxcutsandjobsact, #taxday, #taxcuts, #smallbusinessweek, #economy, #maga, #billionairesfirst, #gopbudget, #goptaxplan, #goptaxbill, #tax, #taxscam, #trumptax

**Immigration**: #DACA, #FamiliesBelongTogether, #Immigration, #MuslimBan, #daca, #familiesbelongtogether, #dreamers, #immigration, #protectdreamers, #dreamactnow, #muslimban, #heretostay, #keepfamiliestogether, #protectthedream, #defenddaca, #immigrants, #familyseparation, #nomuslimbanever, #immigrant, #nobannowall, #borderwall, #refugeeswelcome, #endfamilydetention, #protectfamilies, #refugees

**Abortion**: #ProChoice, #ProLife, #Abortion, #prolife, #abortion, #marchforlife, #prochoice, #theyfeelpain, #bornaliveact, #paincapable, #hr36, #roevwade, #unplanned, #defundpp, #life, #standwithnurses, #endinfanticide, #righttolife, #infanticide, #ppsellsbabyparts

**LGBTQ Rights**: #LGBTQ, #LGBT, #Homophobia, #lgbtq, #lgbt, #equalityact, #pridemonth, #hr5, #nationalcomingoutday, #lgbtqequalityday, #loveislove, #lgbthistorymonth, #transgender, #letkidslearn, #trans, #comingoutday, #marriageequality, #protecttranstroops, #lgbtqhistorymonth,

#defundconversiontherapy, #transban, #otd, #prideinprogress, #nycpride, #protecttranskids, #transrightsarehumanrights, #transdayofremembrance, #loveisthelaw, #rfra, #bathroombill

**Middle-East**: #MiddleEast, #Iran, #Israel, #iran, #israel, #syria, #middleeast, #iraq, #russia, #northkorea, #irandeal, #jordan, #hezbollah, #gaza, #isis, #hamas, #terror, #jihad, #violence, #barbarism, #palestinians, #jewish, #antisemitism, #saudiarabia, #iranian, #lebanon, #turkey, #jerusalem, #iranprotests, #israeli, #freeiran, #sanctions, #supportisrael, #egypt, #terrorism

**Environment**: #ActOnClimate, #ClimateChange, #GreenNewDeal, #climatechange, #actonclimate, #greennewdeal, #climateactionnow, #parisagreement, #climatecrisis, #earthday, #climatefriday, #climate, #climatestrike, #climateaction, #cleanenergy, #climatechangeisreal, #environment, #oceanclimateaction, #cleanair, #climatechangeimpactsme, #cleanwater, #globalwarming, #renewableenergy, #worldenvironmentday, #climateemergency, #peopleoverpolluters, #greenjobs, #climatejustice, #solar, #environmentaljustice, #cleanpowerplan, #todaysclimatefact, #sealevelrise, #bigoil, #climatecrisiscountdown, #stopextinction, #cleanercars, #climatecosts, #cutmethane, #chamberofcarbon, #climatesolutions, #amazonrainforest, #hurricanemaria, #climatesecurityisnationalsecurity, #protectcleanwater, #renewables, #offfossilfuels, #columbiaenergyexchange, #climatesolutionscaucus

# D Additional Visualizations

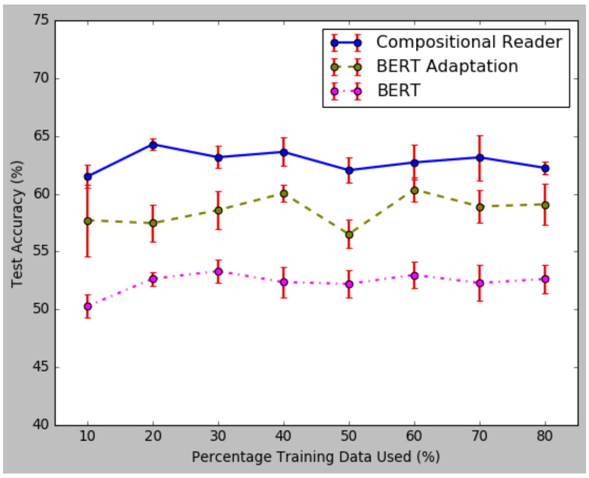


Figure 5: LCV Grade Prediction Task: Training Data Percentage vs Test Accuracy

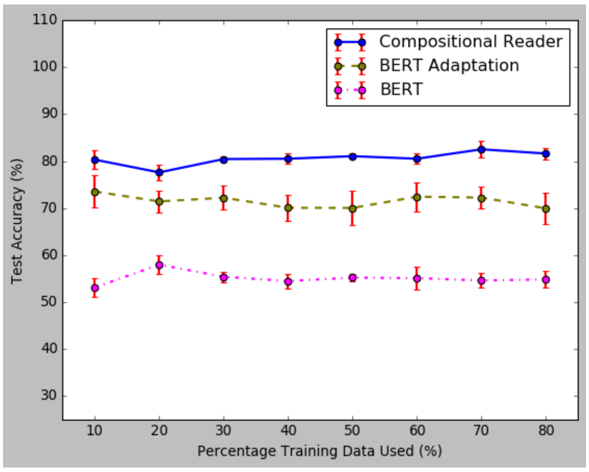


Figure 6: NRA Grade Prediction Task: Training Data Percentage vs Test Accuracy

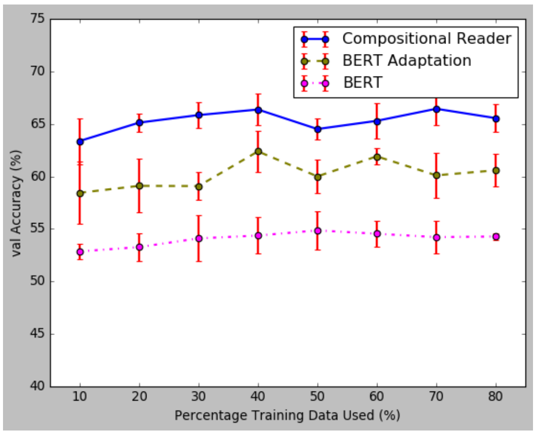


Figure 7: Validation Accuracy on LCV Grade Prediction Task

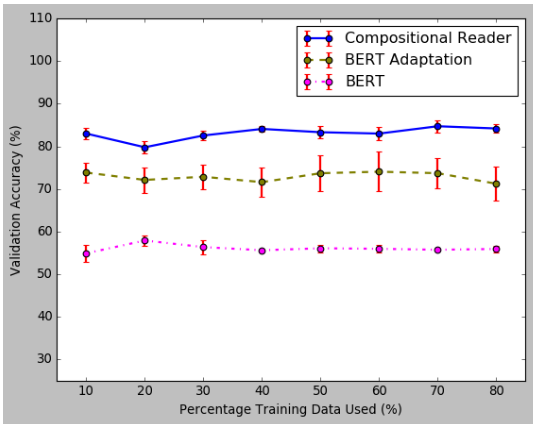


Figure 8: Validation Accuracy on NRA Grade Prediction Task

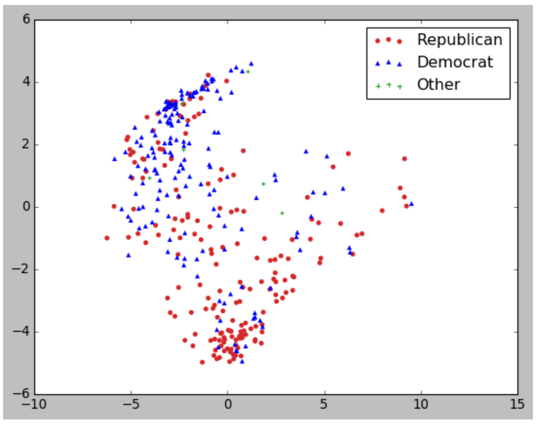


Figure 9: PCA visualization of Republicans vs Democrats on issue *abortion*

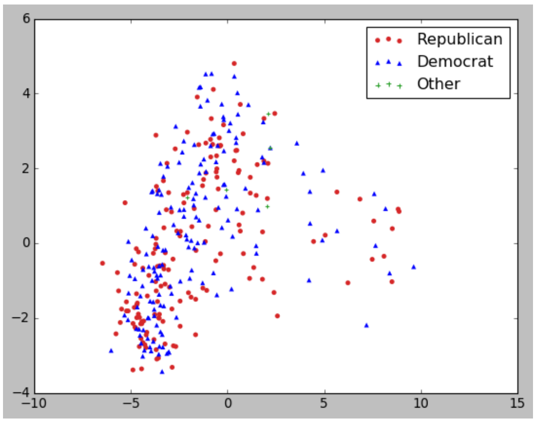


Figure 10: PCA visualization of Republicans vs Democrats on issue *economic-policy*

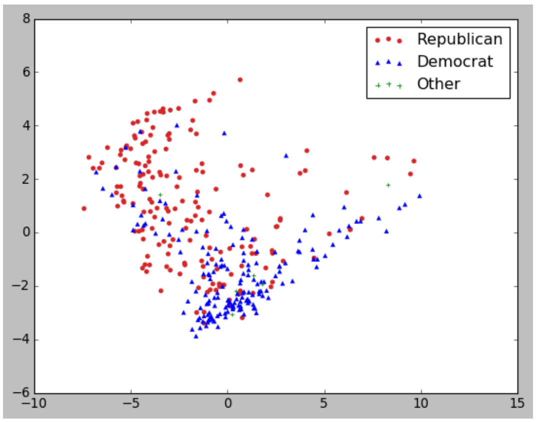


Figure 11: PCA visualization of Republicans vs Democrats on issue *environment*

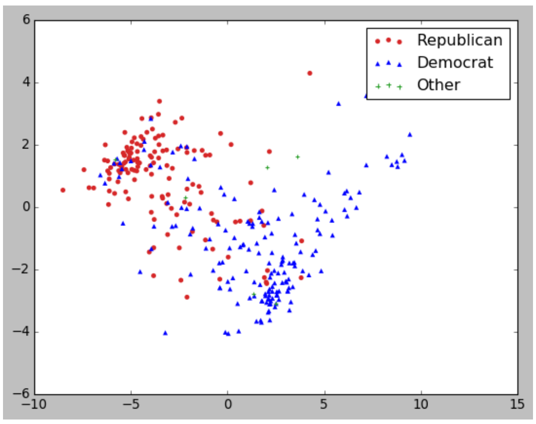


Figure 12: PCA visualization of Republicans vs Democrats on issue *LGBTQ Rights*

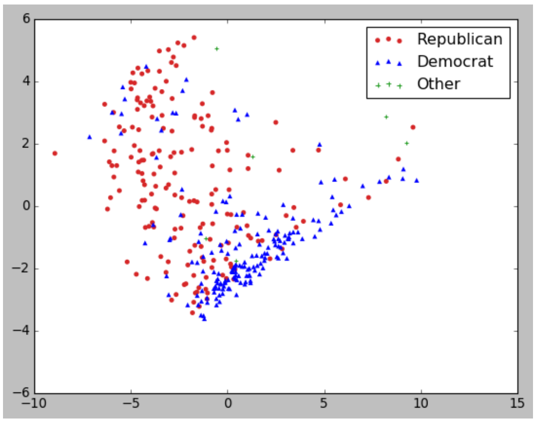


Figure 13: PCA visualization of Republicans vs Democrats on issue *immigration*

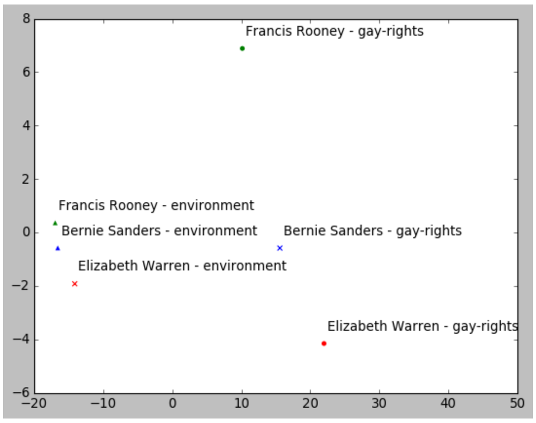


Figure 16: More Comparison of Politician Stances on Issues

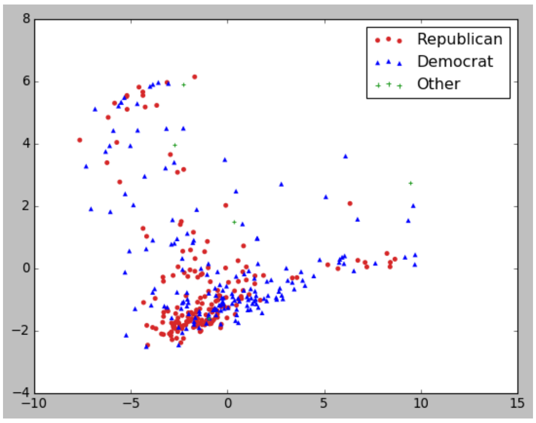


Figure 14: PCA visualization of Republicans vs Democrats on issue *taxes*

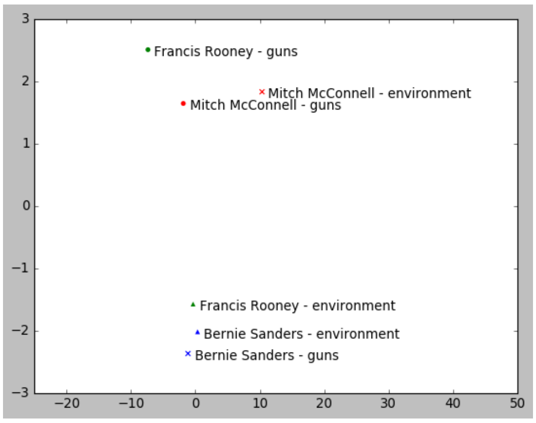


Figure 17: More Comparison of Politician Stances on Issues

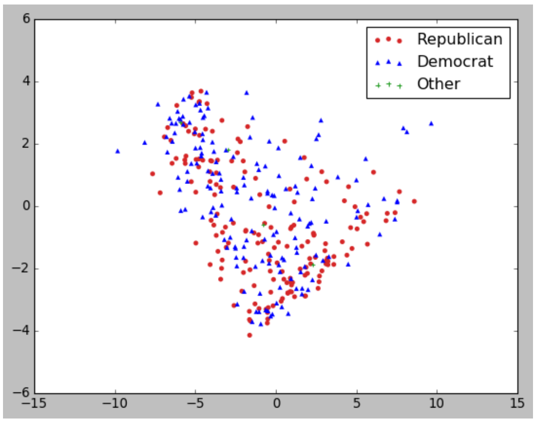


Figure 15: PCA visualization of Republicans vs Democrats on issue *middle-east*

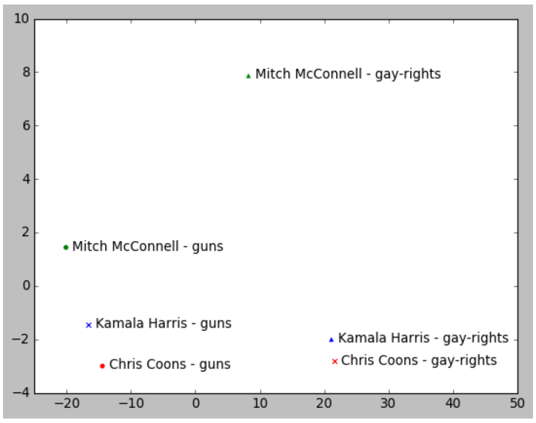


Figure 18: More Comparison of Politician Stances on Issues