

Multi-view Models for Political Ideology Detection of News Articles

Vivek Kulkarni

Department of Computer Science
University of California, Santa Barbara
vvkulkarni@cs.ucsb.edu

Steven Skiena

Department of Computer Science
Stony Brook University
skiena@cs.stonybrook.edu

Junting Ye

Department of Computer Science
Stony Brook University
juyye@cs.stonybrook.edu

William Yang Wang

Department of Computer Science
University of California, Santa Barbara
william@cs.ucsb.edu

Abstract

A news article’s title, content and link structure often reveal its political ideology. However, most existing works on automatic political ideology detection only leverage textual cues. Drawing inspiration from recent advances in neural inference, we propose a novel attention based multi-view model to leverage cues from all of the above views to identify the ideology evinced by a news article. Our model draws on advances in representation learning in natural language processing and network science to capture cues from both textual content and the network structure of news articles. We empirically evaluate our model against a battery of baselines and show that our model outperforms state of the art by 10 percentage points F1 score.

1 Introduction

Many issues covered or discussed by the media and politicians today are so subtle that even word-choice may require one to adopt a particular ideological position (Iyyer et al., 2014). For example, conservatives tend to use the term *tax reform*, while liberals use *tax simplification*. Though objectivity and unbiased reporting remains a cornerstone of professional journalism, several scholars argue that the media displays ideological bias (Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005; Iyyer et al., 2014). Even if one were to argue that such bias may not be reflective of a lack of objectivity, prior research Dardis et al. (2008); Card et al. (2015) note that framing of topics can significantly influence policy.

Since manual detection of political ideology is challenging at a large scale, there has been extensive work on developing computational models for automatically inferring the political ideology of articles, blogs, statements, and congressional speeches (Gentzkow and Shapiro, 2010; Iyyer et al.,

2014; Preoŕiuc-Pietro et al., 2017; Sim et al., 2013). In this paper, we consider the detection of ideological bias at the news article level, in contrast to recent work by Iyyer et al. (2014) who focus on the sentence level or the work of (Preoŕiuc-Pietro et al., 2017) who focus on inferring ideological bias of social media users. Prior research exists on detecting ideological biases of news articles or documents (Gentzkow and Shapiro, 2010; Gerrish and Blei, 2011; Iyyer et al., 2014). However, all of these works generally only model the text of the news article. However, in the online world, news articles do not just contain text but have a rich structure to them. Such an online setting influences the article in subtle ways: (a) choice of the title since this is what is seen in snippet views online (b) links to other news media and sources in the article and (c) the actual textual content itself. Except for the textual content, prior models ignore the rest of these cues. Figure 1 shows an example from The New York Times. Note the presence of hyperlinks in the text, which link to other sources like The Intercept (Figure 1a). We hypothesize that such a link structure is reflective of homophily between news sources sharing similar political ideology – homophily which can be exploited to build improved predictive models (see Figure 1b). Building on this insight, we propose a new model **MVDAM**: *Multi-view document attention model* to detect the ideological bias of news articles by leveraging cues from multiple views: the title, the link structure, and the article content. Specifically, our contributions are:

1. We propose a generic framework **MVDAM** to incorporate multiple views of the news article and show that our model outperforms state of the art by 10 percentage points on the F1 score.
2. We propose a method to estimate the ideological proportions of sources and rank them

by the degree to which they lean towards a particular ideology.

3. Finally, differing from most works, which typically focus on congressional speeches, we conduct ideology detection of news articles by assembling a *large-scale diverse dataset* spanning more than 50 sources.

2 Related Work

Several works study the detection of political ideology through the lens of computational linguistics and natural language processing (Laver et al., 2003; Monroe and Maeda, 2004; Thomas et al., 2006; Lin et al., 2008; Carroll et al., 2009; Ahmed and Xing, 2010; Gentzkow and Shapiro, 2010; Gerrish and Blei, 2011; Sim et al., 2013). Gentzkow and Shapiro (2010) first attempt to rate the ideological leaning of news sources by proposing a measure called “slant index” which captures the degree to which a particular newspaper uses partisan terms or co-allocations. Gerrish and Blei (2011) predict the voting patterns of Congress members based on supervised topic models. Other works use topic models to analyze bias in news articles, blogs, and political speeches (Ahmed and Xing, 2010; Lin et al., 2008). Sim et al. (2013) propose a novel HMM-based model to infer the ideological proportions of the rhetoric used by political candidates in their campaign speeches which relies on a fixed lexicon of bigrams associated with ideologies.

The work that is most closely related to our work is that of Iyyer et al. (2014); Preoțiuc-Pietro et al. (2017). Iyyer et al. (2014) use recurrent neural networks to predict political ideology of congressional debates and articles in the ideological book corpus (IBC) and demonstrate the importance of compositionality in predicting ideology where modifier phrases and punctuality affect the political ideological position. Preoțiuc-Pietro et al. (2017) propose models to infer political ideology of Twitter users based on their everyday language. Most crucially, they also show how to effectively use the relationship between user groups to improve prediction accuracy. Our work draws inspiration from both of these works but differentiates itself from these in the following aspects: We leverage the structure of a news article by noting that an article is just not free-form text, but has a rich structure to it. In particular, we model cues from the title, the inferred network, and the content in a joint generic neural variational inference framework to yield improved

models for this task. Furthermore, differing from Iyyer et al. (2014), we also incorporate attention mechanisms in our model which enables us to inspect which sentences (or words) have the most predictive power as captured by our model. Finally, since we work with news articles (which also contain hyperlinks), naturally our setting is different from all other previous works in general (which mostly focus on congressional debates) and in particular from Iyyer et al. (2014) where only textual content is modeled or Preoțiuc-Pietro et al. (2017) which focuses on social media users.

3 Dataset Construction

News Sources We rely on the data released by ALLSIDES.COM¹ to obtain a list of 59 US-based news sources along with their political ideology ratings: LEFT, CENTER or RIGHT which specify our target label space. While we acknowledge that there is no “perfect” measure of political ideology, ALLSIDES.COM is an apt choice for two main reasons. First, and most importantly the ratings are based on a blind survey, where readers are asked to rate news content without knowing the identity of the news source or the author being rated. This is also precisely the setting in which our proposed computational models operate (where the models have access to the content but are agnostic of the source itself) thus seeking to mirror human judgment closely. Second, these are normalized by ALLSIDES to ensure they closely reflect popular opinion and political diversity present in the United States. These ratings also correlate with independent measurements made by the PEW RESEARCH CENTRE. All these observations suggest that these ratings are fairly robust and generally “reflective of the average judgment of the American People”².

Content Extraction Given the set of news sources selected above, we extract the article content for these news sources. We control for time by obtaining article content over a fixed time-period for all sources. Specifically, we spider several news sources and perform data cleaning. In particular, the spidering component collates the raw HTML of news sources into a storage engine (MongoDB). We track thousands of US based news outlets including country wide popular news sources as well as many local/state news based outlets like the

¹<https://www.allsides.com/media-bias/media-bias-ratings>

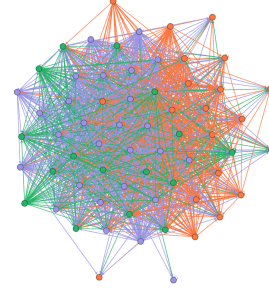
²<https://www.allsides.com/media-bias/about-bias>

A Run of Good Fortune

Around the same time Mr. Pruitt invested in the house in Oklahoma City, he had finished a big business deal that involved Mr. Kelly, Mr. Wagner and a campaign donor who ran a large staffing company.

A baseball player in college, Mr. Pruitt bought an approximately 25 percent stake in the Oklahoma City RedHawks and became the team's managing partner, making him a [highly visible spokesman](#) for the local team. Mr. Wagner also purchased a small stake, and Mr. Kelly's bank provided financing for the deal, as first reported by [The Intercept](#), which also disclosed the bank's loans for one of Mr. Pruitt's suburban Tulsa homes.

(a) A sample news article. Note the presence of hyperlinks to other sources like The Intercept.



(b) Homophily in link structure (viewed in color) of various news sources which can be observed by noting the presence of clusters corresponding to political ideologies. The blue, orange and green clusters correspond to left, right and center leaning sources respectively.

Figure 1: Our proposed framework MVDAM models multiple views of the news article including the content and the link structure. Figure 1a shows a sample article from the New York Times. The presence of such links can provide informative signals for predictive tasks like ideology detection primarily due to homophily (Figure 1b).

Boston Herald³. However, in this paper, we consider only the 59 US news sources for which we can derive ground truth labels for political ideology. For each of the news sources considered, we extract the title, the cleaned pre-processed content, and the hyperlinks within the article that reveal the network structure. The label for each article is the label assigned to its source as obtained from ALL-SIDES. We choose a random sample of 120,000 articles and create 3 independent splits for training (100,000), validation (10,000) and test (10,000) with a roughly balanced label distribution.⁴

Data Pre-processing and Cleaning Since the labels were derived from the source, we are careful to remove any systematic features in each article which are trivially reflective of the source, since that would result in over-fitting. In particular we perform the following operations: (a) **Remove source link mentions** When modeling the link structure of an article, we explicitly remove any link to the source itself. Second, we also explicitly remove any systematic link structures in articles that are source specific. In particular, some sources may always have links to other domains (like their own franchisees or social media sites). These links are removed explicitly by noting their high frequency. (b) **Remove headers, footers, advertisements** News sources systematically introduce footers, and advertisements which we remove explicitly. For example, every

article of the The Daily Beast has the following footer You can subscribe to the Daily Beast here which we filter out.

4 Models and Methods

Problem Formulation Given $\mathbf{X} = \{\mathbf{X}_{title}, \mathbf{X}_{net}, \mathbf{X}_{content}\}$ which represents a set of multi-modal features of news articles and a label set $\mathbf{Y} = \{\text{LEFT}, \text{CENTER}, \text{RIGHT}\}$, we would like to model $\Pr(\mathbf{Y}|\mathbf{X})$.

Overview of MVDAM We consider a Bayesian approach with stochastic attention units to effectively model textual cues. Bayesian approaches with stochastic attention have been noted to be quite effective at modeling ambiguity as well as avoiding over-fitting scenarios especially in the case of small training data sets (Miao et al., 2016). In particular, we assume a latent representation \mathbf{h} learned from the multiple modalities in \mathbf{X} which is then mapped to the label space \mathbf{Y} . In the most general setting, instead of learning a deterministic encoding \mathbf{h} given \mathbf{X} , we posit a latent distribution over the hidden representation \mathbf{h} , $\Pr(\mathbf{h}|\mathbf{X})$ to model the overall document where $\Pr(\mathbf{h}|\mathbf{X})$ is parameterized by a diagonal Gaussian distribution $\mathcal{N}(\mathbf{h}|\mu(\mathbf{X}), \sigma^2(\mathbf{X}))$.

Specifically, consider the distribution $\Pr(\mathbf{Y}|\mathbf{X})$ which can be written as follows:

$$\Pr(\mathbf{Y}|\mathbf{X}) = \sum_{\mathbf{h}} \Pr(\mathbf{Y}|\mathbf{h}) \Pr(\mathbf{h}|\mathbf{X}) \quad (1)$$

As noted by Miao et al. (2016), computing the exact posterior is in general intractable. Therefore, we posit a variational distribution $q_{\phi}(\mathbf{h})$ and maximize

³This is a part of an ongoing project called MediaRank. More details can be found at <http://media-rank.com>

⁴Note that we do not restrict the articles to be strictly political since even articles on other topics like health and sports can be reflective of political ideology (Hoberman, 1977).

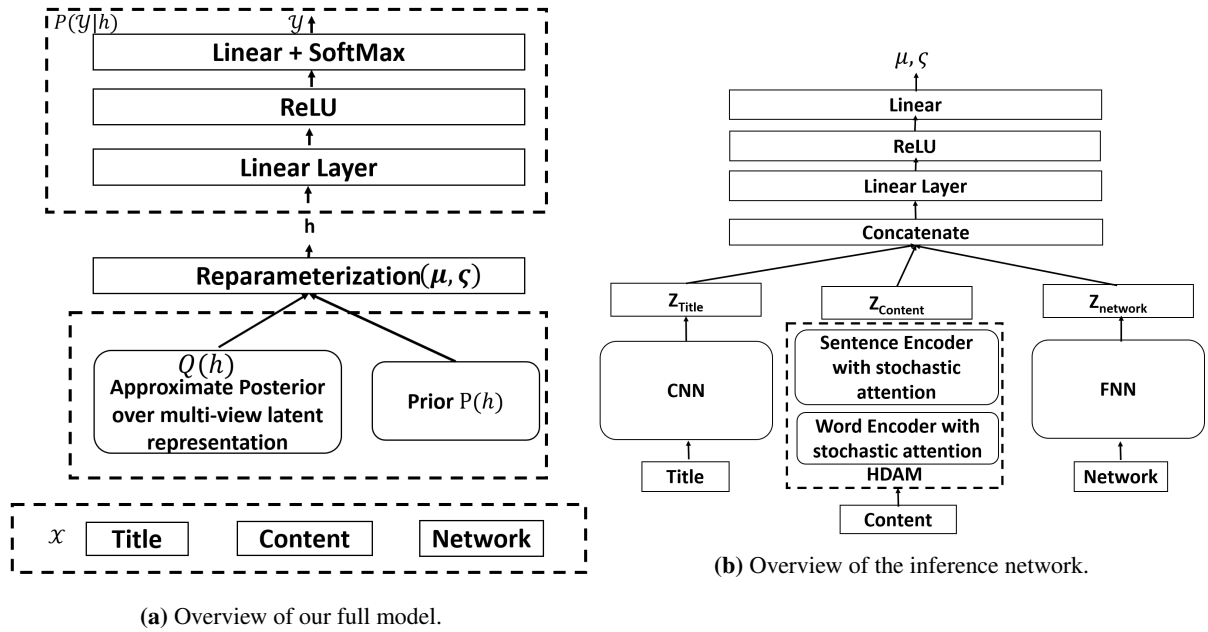


Figure 2: A broad overview of our MVDAM model depicting the three major components: a discriminator, an inference network and a prior and captures cues from multiple views of the news article. As noted by Miao et al. (2016) we use stochastic attention units which are shown to model ambiguity better. We thus train the model end-to-end using neural variational inference.

the evidence lower bound $\mathcal{L} \leq \Pr(\mathbf{y}|\mathbf{X})$ namely,

$$\mathcal{L} = \mathbb{E}_{q_\phi(\mathbf{h})}[p(\mathbf{Y}|\mathbf{h})] - D_{KL}(q_\phi(\mathbf{h})||p(\mathbf{h}|\mathbf{X})), \quad (2)$$

where $p(\mathbf{Y}|\mathbf{h})$ denotes a probability distribution over \mathbf{Y} given the latent representation \mathbf{h} , and $p(\mathbf{h}|\mathbf{X})$ denotes the probability distribution over \mathbf{h} conditioned on \mathbf{X} .

Equation 2 can be interpreted as consisting of three components, each of which can be modeled separately: (a) **Discriminator** $p(\mathbf{Y}|\mathbf{h})$ can be viewed as a discriminator given the hidden representation \mathbf{h} . Maximizing the first term is thus equivalent to minimizing the cross-entropy loss between the model’s prediction and true labels. (b) The second term, the KL Divergence term consists of two components: (1) **Approximate Posterior** The term $q_\phi(\mathbf{h})$ also known as the approximate posterior parameterizes the latent distribution which encodes the multi-modal features \mathbf{X} of a document. (2) **Prior** The term $p(\mathbf{h}|\mathbf{X})$ can be viewed as a prior which can be uninformative (a standard Gaussian prior in the most general case, or any other prior model based on other features). We now discuss how we model each of these components in detail.

4.1 Discriminator

We use a simple feed-forward network with a linear layer that accepts as input the latent hidden representation of \mathbf{X} , followed by a ReLU for non-

linearity followed by a linear layer and a final softmax layer to model this component.

4.2 Approximate Posterior

Here we model the approximate posterior $q_\phi(\mathbf{h})$ by an inference network shown succinctly in Figure 2b. The inference network takes as input the features \mathbf{X} and learns a corresponding hidden representation \mathbf{h} . More specifically, it outputs two components: (μ, ζ) corresponding to the mean and log-variance of the gaussian parametrizing the hidden representation \mathbf{h} . We model this using a “multi-view” network which incorporates hidden representations learned from multiple modalities into a joint representation. Specifically, given d -dimensional hidden representations corresponding to multiple modalities z_{title} , $z_{network}$, and $z_{content}$ the model first concatenates these representations into a single $3d$ -dimensional representation z_{concat} which is then input through a 2-layer feed-forward network to output a d -dimensional mean vector μ and a d -dimensional log-variance vector ζ that parameterizes the latent distribution governing \mathbf{h} . We now discuss the models used for capturing each view.

4.2.1 Modeling the Title

We learn a latent representation of the title of a article by using a convolutional network. Convolutional networks have been shown to be very effec-

tive for modeling short sentences like titles of news articles. In particular, we use the same architecture proposed by (Kim, 2014). The input words of the title are mapped to word embeddings and concatenated and passed through convolutional filters of varying window sizes. This is then followed by a max-over-time pooling (Collobert et al., 2011). The outputs of this layer are input to a fully connected layer of dimension d with drop-out which outputs z_{title} , the latent representation of the title.

4.2.2 Modeling the Network Structure of articles

Capturing the network structure of article consists of two steps: (a) Learning a network representation of each source based on its social graph G . (b) Using the learned representation of each source to capture the link structure of a particular article.

We use a state-of-the-art network representation learning algorithm to learn representations of nodes in a social network. In particular, we use Node2Vec (Grover and Leskovec, 2016), which learns a d -dimensional representation of each source given the hyperlink structure graph G . Node2Vec seeks to maximize the log likelihood of observing the neighborhood of a node $\mathcal{N}(u)$, given the node u . Let \mathbf{F} be a matrix of size (V, d) where $\mathbf{F}(u)$ represents the embedding of node u . We then maximize the following likelihood function $\max_{\mathbf{F}} \sum_u \log \Pr(\mathcal{N}(u)|u)$. We model the above likelihood similar to the Skip-gram architecture (Mikolov et al., 2013) by assuming that the likelihood of observing a node $v \in \mathcal{N}(u)$ is conditionally independent of any other node in the neighborhood given u . That is $\log \Pr(\mathcal{N}(u)|u) = \sum_{v \in \mathcal{N}(u)} \log \Pr(v|u)$. We then model $\Pr(v|u) = \frac{e^{\mathbf{F}(u) \cdot \mathbf{F}(v)}}{\sum_v e^{\mathbf{F}(u) \cdot \mathbf{F}(v)}}$. Having fully specified the log likelihood function, we can now optimize it using stochastic gradient ascent.

Having learned the embedding matrix \mathbf{F} for each source node, we now model the link structure of any given article \mathcal{A} simply by the average of the network embedding representations for each link l referenced in \mathcal{A} . In particular, we compute $z_{network}$ as: $z_{network} = \frac{1}{|\mathcal{A}|} \sum_{l \in \mathcal{A}} \mathbf{F}(l)$.

4.2.3 Modeling the Content of articles

To model the content of an article, we use a hierarchical approach with attention. In particular, we compute attention at both levels: (a) words and (b) sentences. We closely follow the approach by

(Yang et al., 2016) which learns a latent representation of a document d using both word and sentence attention models.

We model the article \mathcal{A} hierarchically, by first representing each sentence i with a hidden representation s_i . We model the fact that not all words contribute equally in the sentence through a word level attention mechanism. We then learn the representation of the article \mathcal{A} by composing these individual sentence level representations with a sentence level attention mechanism.

Learning sentence representations We first map each word to its embedding matrix through a lookup embedding matrix \mathbf{W} . We then learn a hidden representation of the given sentence h_{it} centered around word w_i by embedding the sentence through a bi-directional GRU as described by (Bahdanau et al., 2014). Since not all words contribute equally to the representation of the sentence, we introduce a word level attention mechanism which attempts to extract relevant words that contribute to the meaning of the sentence. Specifically we learn a word level attention matrix \mathbf{W}_w as follows $\alpha_i \propto \exp(\mathbf{W}_w h_{it} + b_w)$, $s_i = \sum_t \alpha_i h_{it}$ where s_i is the latent representation of the sentence i .

Composing sentence representations We follow a similar method to learn a latent representation of an article. Given the embedding s_i of each sentence in the article, we learn a hidden representation of the given sentence h_i centered around s_i by embedding the list of sentences through a bi-directional GRU as described by (Bahdanau et al., 2014). Once again, since not all sentences contribute equally to the representation of the article, we introduce a sentence level attention mechanism which attempts to extract relevant sentences that contribute to the meaning of the article. Specifically we learn the weights of a sentence level attention matrix \mathbf{W}_s as $\alpha_s \propto \exp(\mathbf{W}_s h_s + b_s)$, $z_{content} = \sum_s \alpha_s h_s$, where $z_{content}$ is the latent representation of the article. In this case we let the hidden representation of the sentence be a stochastic representation similar to the work by (Miao et al., 2016) and use the Gaussian re-parameterization trick to enable training via end-to-end gradient based methods⁵. Such techniques have been shown to be useful in

⁵Using deterministic sentence representations is a special case.

modeling ambiguity and also generalize well to small training datasets (Miao et al., 2016).

4.3 Prior

The prior models $p(\mathbf{h}|\mathbf{X})$ in Equation 2. Note that our proposed framework is general and can be used to incorporate a variety of priors. Here, we assume the prior is drawn from a Gaussian distribution with diagonal co-variances. The KL Divergence term in Equation 2 can thus be analytically computed. In particular, the KL Divergence between two K dimensional Gaussian distributions \mathbf{A}, \mathbf{B} with means μ_A, μ_B and diagonal co-variances κ_A, κ_B is:

$$D_{KL}(\mathbf{A}, \mathbf{B}) = -\frac{1}{2} \sum_{j=1}^{j=K} \left(1 + \log \frac{\kappa_{Aj}}{\kappa_{Bj}} - \frac{\kappa_{Aj}}{\kappa_{Bj}} - (\mu_{Aj} - \mu_{Bj})^2 / \kappa_{Bj} \right) \quad (3)$$

Parameter Estimation Having described precisely, the models for each of the components in Equation 2, we can reformulate the maximization of the variational lower bound to the following loss function on the set of all learn-able model parameters θ : $\mathcal{J}(\theta)$ as follows:

$$\mathcal{J}(\theta) = \text{NLL}(\mathbf{y}|\mathbf{X}) + \lambda D_{KL}(q(\mathbf{h})||p(\mathbf{h}|\mathbf{X})), \quad (4)$$

where NLL is the negative log likelihood loss computed between the predicted label and the true label, and λ is a hyper-parameter that controls the amount of regularization offered by the KL Divergence term. We use ADADELTA to minimize this loss function.

5 Experiments

We evaluate our model against several competitive baselines which model only a single view to place our model in context:

1. **Chance Baseline** We consider a simple baseline that returns a draw from the label distribution as the prediction.
2. **Logistic Regression LR (Title)** We consider a bag of words classifier using Logistic Regression that can capture linear relationships in the feature space and use the words of the title as the feature set.
3. **CNN (Title)** We consider a convolutional net classifier based on exactly the same architecture as (Kim, 2014) which uses the title of the news article. Convolutional Nets have been

shown to be extremely effective at classifying short pieces of text and can capture non-linearities in the feature space (Kim, 2014).

4. **FNN (Network)** We also consider a simple fully-connected feed forward neural network using only the network features to characterize the predictive power of the network alone.
5. **HDAM Model (Content)** We use the state of the art hierarchical document attention model proposed by (Yang et al., 2016) that models the content of the article using both word and sentence level attention mechanisms.

We consider three different flavors of our proposed model which differ in the subset of modalities used (a) **Title and Network** (b) **Title and Content**, and (c) **Full model**: Title, Network, and Content. We train all of our models and the baselines on the training data set choosing all hyper-parameter using the validation set. We report the performance of all models on the held-out test set.

Experimental Settings We set the embedding latent dimension captured by each view to be 128 including the final latent representation obtained by fusing multiple modalities. In case of the CNN’s, we consider three convolutional filters of window sizes 3, 4, 5 each yielding a 100 dimensional feature map followed by max-over time pooling which is then passed through a fully connected layer to yield the output. In all the neural models, we used AdaDelta with an initial learning rate of 1.0 to learn the parameters of the model via back-propagation.

Model	Views	P	R	F1
CHANCE	–	34.53	34.59	34.53
LR	Title	59.53	59.42	59.12
CNN	Title	59.26	59.40	59.24
FNN	Network	68.28	56.54	55.10
HDAM	Content	69.85	68.72	68.92
MVDAM	Title, Net- work	69.87	69.71	69.66
MVDAM	Title, Content	70.84	70.19	69.54
MVDAM	Title, Net- work, Content	80.10	79.56	79.67

Table 1: Precision, Recall, and F1 scores of our model MVDAM on the test set compared with several baselines. All flavors of our model significantly outperform baselines and yield state of the art performance.

5.1 Results and Analysis

Quantitative Results Table 1 shows the results of the evaluation. First note that the logistic regression classifier and the CNN model using the Title outperforms the CHANCE classifier significantly

(F1: **59.12**, **59.24** vs **34.53**). Second, only modeling the network structure yields a F1 of **55.10** but still significantly better than the chance baseline. This confirms our intuition that modeling the network structure can be useful in prediction of ideology. Third, note that modeling the content (HDAM) significantly outperforms all previous baselines (F1:**68.92**). This suggests that content cues can be very strong indicators of ideology. Finally, all flavors of our model outperform the baselines. Specifically, observe that incorporating the network cues outperforms all uni-modal models that only model either the title, the network, or the content. It is also worth noting that without the network, only the title and the content show only a small improvement over the best performing baseline (**69.54** vs **68.92**) suggesting that the network yields distinctive cues from both the title, and the content. Finally, the best performing model effectively uses all three modalities to yield a F1 score of **79.67** outperforming the state of the art baseline by 10 percentage points. Altogether our results suggest the superiority of our model over competitive baselines. In order to obtain deeper insights into our model, we also perform a qualitative analysis of our model's predictions.

Visualizing Attention Scores Figure 3 shows a visualization of sentences based on their attention scores. Note that for a left leaning article (see Figure 3a), the model focuses on sentences involving *gun-control*, *feminists*, and *transgender*. In contrast, a visualization of sentence attention scores for an article which the model predicted as “right-leaning” (see Figure 3b)) reveals a focus on words like *god*, *religion* etc. These observations qualitatively suggest that the model is able to effectively pick up on content cues present in the article. By examining the distribution over the sentence indices corresponding to the maximum attention scores, we noted that only in about half the instances, the model focuses its greatest attention on the beginning of the article suggesting that the ability to selectively focus on sentences in the news article contributes to the superior performance.

Challenging Cases In Table 2, we highlight some of the challenges of our model. In particular, our model finds it quite challenging to identify the political ideology of the source for articles that are non-political and related to global events, or

entertainment. Examples include instances like *Tourist dies hiking in Australia* *Outback heat* or *Juan Williams makes the 'case for Oprah'*. We also note that articles with “click-bait” titles like *We are all Just Overclocked Chimpanzees* are not necessarily discriminative of the underlying ideology. In summary, while our proposed model significantly advances the state of art, it also suggests scope for further improvement especially in identifying political ideologies of articles in topics like Entertainment or Sports. For example, prior research suggests that engagement in particular sports is correlated with the political leanings (Hoberman, 1977) which suggest that improved models might need to capture deeper linguistic and contextual cues.

Ideological Proportions of News Sources Finally, we compute the expected proportion of an ideology in a given source based on the probability estimates output by our model for the various articles. While one might expect that the expected degree of “left-ness” (or “right-ness”) for a given source can easily be computed by taking a simple mean of the prediction probability for the given ideology over all articles belonging to the source, such an approach can be in-accurate because the probability estimates output by the model are not necessarily calibrated and therefore cannot be interpreted as a confidence value. We therefore use isotonic regression to calibrate the probability scores output by the model. Having calibrated the probability scores, we now compute the degree to which a particular news source leans toward an ideology by simply computing the mean output score over all articles corresponding to the source. Table 3 shows the top 10 sources ranked according to their proportions for each ideology. We note that sources like CNN, Buzz Feed, SF Chronicle are considered more left-leaning than the Washington Post. Similarly, we note that NPR and Reuters are considered to be the most center-aligned while Breitbart, Infowars and Blaze are considered to be most right-aligned by our model. These observations are moderately aligned with survey results that place news sources on the ideology spectrum based on the political beliefs of their consumers⁶.

⁶http://www.journalism.org/2014/10/21/political-polarization-media-habits/pj.14-10-21_mediapolarization-08/

There are so many go to GOP scapegoats it would be impossible to name them all here .I 'msure many of you have your favorites Black Lives Matter the EPA feminists climatologists Islam income taxes Michael Moore Hillary Clinton and anybody who ever says the words "gun control" .If you can name it they can blame it .In no particular order here are eight of the GOP 's favorite scapegoats .Transgender people It 's no secret that the GOP is the antithesis of queer friendly and the latest arm of this prejudice is aimed at the transgender community .Widespread attention to the issue came in 2016 when North

(a) Sample attention on sentences for a Left aligned article.

Many people approach this occasion with the best of intentions -to slow down be more patient and pray .But in our busy fast paced world taking our foot off the gas and finding time to pray can often be difficult .As an editor of books by such authors as Pope Benedict XVI and Pope Francis and as the author of "Life Everlasting Catholic Devotions and Mysteries for the Everday Seeker " Isometimes find it hard to pray too .But some years back a friend of mine turned me on to a 400 year old spiritual exercise that has revolutionized my life ultimately helping me feel less anxious and stressed .It 's called the examen a Christian mindfulness tool developed by 16th century mystic St .Its objective is to bring us closer to God .It 's a perfect exercise not just for people preparing for Easter but for all who believe in God and need a little help shifting focus .The examen is a series of short reflections and questions that asks us first what is drawing us near God and second what is drawing us away from God .What fears and worries are blocking

(b) Sample attention on sentences for a Right aligned article.

Figure 3: Visualization of attention on different sentences on two sample articles from the Left and Right aligned sources respectively. Note the different focus based in the ideology reflected by the highlighted words.

Article Title	Source Label	Predicted Label
Juan Williams Makes the 'Case for Oprah'	Right	Left
Tourist dies hiking in Australia Outback heat	Right	Left
Back From China, UCLA Basketball Players Plagued by Father	Right	Left
Democrat Ralph Northam Elected Governor of Virginia	Right	Left
South Africa blighted by racially charged farm murders	Right	Left
Lawsuit: Stripper punched man, knocked out his front tooth	Left	Right
Heres How to Keep Fake News Off Twitter	Left	Right
We Are All Just Overclocked Chimpanzee	Left	Right
Curious Arctic Fox Pups Destroy Hidden Camera In The Most Adorable Way	Left	Right
I am American, Jewish, and banned from Israel for my activism	Left	Right

Table 2: Few failure cases of our model illustrating what our model finds challenging. Articles with "click-baity" titles are not necessarily very discriminative of the ideology. Similarly, articles that are *non-political* and related to global events or entertainment are quite challenging.

Rank	Source	Rank	Source	Rank	Source
1	CNN	1	NPR	1	Breitbart
2	BuzzFeed	2	Reuters	2	Infowars
3	SF Chronicle	3	USA Today	3	Blaze
4	CBS News	4	BBC	4	Fox News
5	BoingBoing	5	CNBC	5	KSL
6	Mother Jones	6	Chicago Tribune	6	Townhall
7	Think Progress	7	Business Insider	7	CBN
8	The Atlantic	8	Forbes	8	ConservativeHQ
9	The Washington Post	9	APR	9	NewsMax
10	Rolling Stone	10	The Wall Street Journal	10	DailyWire

(a) Left aligned

(b) Centre aligned

(c) Right aligned

Table 3: A Top 10 ranking of Ideological sources as obtained by our model which correlate moderately with external surveys.

6 Conclusion

We proposed a model to leverage cues from multiple views in the predictive task of detecting political ideology of news articles. We show that incorporating cues from the title, the link structure and the content significantly beats state of the art. Finally, using the predicted probabilities of our model, we draw on methods for probability cali-

bration to rank news sources by their ideological proportions which moderately correlates with existing surveys on the ideological placement of news sources. To conclude, our proposed framework effectively leverages cues from multiple views to yield state of the art interpret-able performance and sets the stage for future work which can easily incorporate other modalities like audio, video and images.

Acknowledgments

We thank the anonymous reviewers for their comments. This research was supported in part by DARPA Grant D18AP00044 funded under the DARPA YFA program. This work was also partially supported by NSF grants DBI-1355990 and IIS-1546113. The authors are solely responsible for the contents of the paper, and the opinions expressed in this publication do not reflect those of the funding agencies.

References

- Amr Ahmed and Eric P Xing. 2010. Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1140–1150. Association for Computational Linguistics.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Dallas Card, Amber E Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 438–444.
- Royce Carroll, Jeffrey B Lewis, James Lo, Keith T Poole, and Howard Rosenthal. 2009. Measuring bias and uncertainty in dw-nominate ideal point estimates via the parametric bootstrap. *Political Analysis*, 17(3):261–275.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537.
- Frank E Dardis, Frank R Baumgartner, Amber E Boydston, Suzanna De Boef, and Fuyuan Shen. 2008. Media framing of capital punishment and its impact on individuals’ cognitive responses. *Mass Communication & Society*, 11(2):115–140.
- Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.
- Sean Gerrish and David M Blei. 2011. Predicting legislative roll calls from text. In *Proceedings of the 28th international conference on machine learning (icml-11)*, pages 489–496.
- Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.
- Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- John M Hoberman. 1977. Sport and political ideology. *Journal of sport and social issues*, 1(2):80–114.
- Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1113–1122.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331.
- Wei-Hao Lin, Eric Xing, and Alexander Hauptmann. 2008. A joint topic and perspective model for ideological discourse. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 17–32. Springer.
- Yishu Miao, Lei Yu, and Phil Blunsom. 2016. Neural variational inference for text processing. In *International Conference on Machine Learning*, pages 1727–1736.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Burt L Monroe and Ko Maeda. 2004. Talks cheap: Text-based estimation of rhetorical ideal-points.
- Daniel Preoțiuc-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)*, volume 1, pages 729–740.
- Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 91–101.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489.