Political Ideology Detection Using Recursive Neural Networks

Mohit Iyyer[1], **Peter Enns**[2], **Jordan Boyd-Graber**[3,4], **Philip Resnik**[2,4]
[1]Computer Science, [2]Linguistics, [3]iSchool, and [4]UMIACS
University of Maryland
{miyyer,peter,jbg}@umiacs.umd.edu, resnik@umd.edu

Abstract

An individual's words often reveal their political ideology. Existing automated techniques to identify ideology from text focus on bags of words or wordlists, ignoring syntax. Taking inspiration from recent work in sentiment analysis that successfully models the compositional aspect of language, we apply a recursive neural network (RNN) framework to the task of identifying the political position evinced by a sentence. To show the importance of modeling subsentential elements, we crowdsource political annotations at a phrase and sentence level. Our model outperforms existing models on our newly annotated dataset and an existing dataset.

1 Introduction

Many of the issues discussed by politicians and the media are so nuanced that even word choice entails choosing an ideological position. For example, what liberals call the "estate tax" conservatives call the "death tax"; there are no ideologically neutral alternatives (Lakoff, 2002). While objectivity remains an important principle of journalistic professionalism, scholars and watchdog groups claim that the media are biased (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Niven, 2003), backing up their assertions by publishing examples of obviously biased articles on their websites. Whether or not it reflects an underlying lack of objectivity, quantitative changes in the popular framing of an issue over time—favoring one ideologically-based position over another—can have a substantial effect on the evolution of policy (Dardis et al., 2008).

Manually identifying ideological bias in political text, especially in the age of big data, is an impractical and expensive process. Moreover, bias may be localized to a small portion of a document, undetectable by coarse-grained methods. In this paper, we examine the problem of detecting ideological bias on the sentence level. We say a sentence contains *ideological bias* if its author's political position (here *liberal* or *conservative*, in the sense of U.S. politics) is evident from the text.

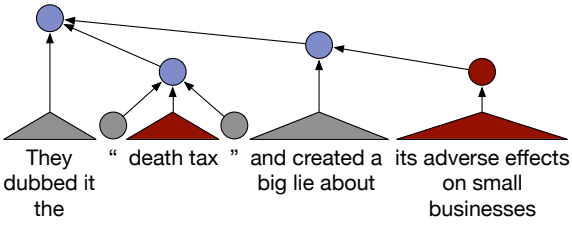

Figure 1: An example of compositionality in ideological bias detection (red → conservative, blue → liberal, gray → neutral) in which modifier phrases and punctuation cause polarity switches at higher levels of the parse tree.

Ideological bias is difficult to detect, even for humans—the task relies not only on political knowledge but also on the annotator's ability to pick up on subtle elements of language use. For example, the sentence in Figure 1 includes phrases typically associated with conservatives, such as "small businesses" and "death tax". When we take more of the structure into account, however, we find that scare quotes and a negative propositional attitude (*a lie about X*) yield an evident liberal bias.

Existing approaches toward bias detection have not gone far beyond "bag of words" classifiers, thus ignoring richer linguistic context of this kind and often operating at the level of whole documents. In contrast, recent work in sentiment analysis has used deep learning to discover compositional effects (Socher et al., 2011b; Socher et al., 2013b).

Building from those insights, we introduce a recursive neural network (RNN) to detect ideological bias on the sentence level. This model requires

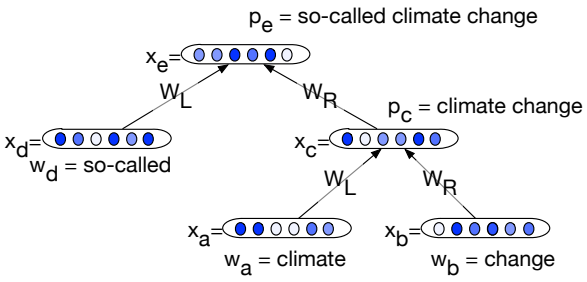

Figure 2: An example RNN for the phrase "so-called climate change". Two d-dimensional word vectors (here, $d = 6$) are composed to generate a phrase vector of the same dimensionality, which can then be recursively used to generate vectors at higher-level nodes.

richer data than currently available, so we develop a new political ideology dataset annotated at the phrase level. With this new dataset we show that RNNs not only label sentences well but also improve further when given additional phrase-level annotations. RNNs are quantitatively more effective than existing methods that use syntactic and semantic features separately, and we also illustrate how our model correctly identifies ideological bias in complex syntactic constructions.

2 Recursive Neural Networks

Recursive neural networks (RNNs) are machine learning models that capture syntactic and semantic composition. They have achieved state-of-the-art performance on a variety of sentence-level NLP tasks, including sentiment analysis, paraphrase detection, and parsing (Socher et al., 2011a; Hermann and Blunsom, 2013). RNN models represent a shift from previous research on ideological bias detection in that they do not rely on hand-made lexicons, dictionaries, or rule sets. In this section, we describe a supervised RNN model for bias detection and highlight differences from previous work in training procedure and initialization.

2.1 Model Description

By taking into account the hierarchical nature of language, RNNs can model *semantic composition*, which is the principle that a phrase's meaning is a combination of the meaning of the words within that phrase and the syntax that combines those words. While semantic composition does not apply universally (e.g., sarcasm and idioms), most language follows this principle. Since most ideological bias becomes identifiable only at higher levels of sentence trees (as verified by our annotation, Figure 4), models relying primarily on word-level distributional statistics are not desirable for our problem.

The basic idea behind the standard RNN model is that each word w in a sentence is associated with a vector representation $x_w \in \mathbb{R}^d$. Based on a parse tree, these words form phrases p (Figure 2). Each of these phrases also has an associated vector $x_p \in \mathbb{R}^d$ of the same dimension as the word vectors. These phrase vectors should represent the meaning of the phrases composed of individual words. As phrases themselves merge into complete sentences, the underlying vector representation is trained to retain the sentence's whole meaning.

The challenge is to describe how vectors combine to form complete representations. If two words w_a and w_b merge to form phrase p, we posit that the phrase-level vector is

$$x_p = f(W_L \cdot x_a + W_R \cdot x_b + b_1), \qquad (1)$$

where W_L and W_R are $d \times d$ left and right composition matrices shared across all nodes in the tree, b_1 is a bias term, and f is a nonlinear activation function such as $\tanh$. The word-level vectors x_a and x_b come from a $d \times V$ dimensional word embedding matrix W_e, where V is the size of the vocabulary.

We are interested in learning representations that can distinguish political polarities given labeled data. If an element of this vector space, x_d, represents a sentence with liberal bias, its vector should be distinct from the vector x_r of a conservative-leaning sentence.

Supervised RNNs achieve this distinction by applying a regression that takes the node's vector x_p as input and produces a prediction $\hat{y}_p$. This is a softmax layer

$$\hat{y}_p = \text{softmax}(W_{cat} \cdot x_p + b_2), \qquad (2)$$

where the softmax function is

$$\text{softmax}(q) = \frac{\exp q}{\sum_{j=1}^{k} \exp q_j} \qquad (3)$$

and W_{cat} is a $k \times d$ matrix for a dataset with k-dimensional labels.

We want the predictions of the softmax layer to match our annotated data; the discrepancy between categorical predictions and annotations is measured

through the cross-entropy loss. We optimize the model parameters to minimize the cross-entropy loss over all sentences in the corpus. The cross-entropy loss of a single sentence is the sum over the true labels y_i in the sentence,

$$\ell(\hat{y}_s) = \sum_{p=1}^{k} y_p * log(\hat{y}_p). \quad (4)$$

This induces a supervised objective function over all sentences: a regularized sum over all node losses normalized by the number of nodes N in the training set,

$$C = \frac{1}{N} \sum_{i}^{N} \ell(pred_i) + \frac{\lambda}{2} \|\theta\|^2. \quad (5)$$

We use L-BFGS with parameter averaging (Hashimoto et al., 2013) to optimize the model parameters $\theta = (W_L, W_R, W_{cat}, W_e, b_1, b_2)$. The gradient of the objective, shown in Eq. (6), is computed using backpropagation through structure (Goller and Kuchler, 1996),

$$\frac{\partial C}{\partial \theta} = \frac{1}{N} \sum_{i}^{N} \frac{\partial \ell(\hat{y}_i)}{\partial \theta} + \lambda\theta. \quad (6)$$

2.2 Initialization

When initializing our model, we have two choices: we can initialize all of our parameters randomly or provide the model some prior knowledge. As we see in Section 4, these choices have a significant effect on final performance.

Random The most straightforward choice is to initialize the word embedding matrix W_e and composition matrices W_L and W_R randomly such that without any training, representations for words and phrases are arbitrarily projected into the vector space.

word2vec The other alternative is to initialize the word embedding matrix W_e with values that reflect the meanings of the associated word types. This improves the performance of RNN models over random initializations (Collobert and Weston, 2008; Socher et al., 2011a). We initialize our model with 300-dimensional *word2vec* toolkit vectors generated by a continuous skip-gram model trained on around 100 billion words from the Google News corpus (Mikolov et al., 2013).

The word2vec embeddings have linear relationships (e.g., the closest vectors to the average of "green" and "energy" include phrases such as "renewable energy", "eco-friendly", and "efficient lightbulbs"). To preserve these relationships as phrases are formed in our sentences, we initialize our left and right composition matrices such that parent vector p is computed by taking the average of children a and b ($W_L = W_R = 0.5\mathbb{I}_{d\times d}$). This initialization of the composition matrices has previously been effective for parsing (Socher et al., 2013a).

3 Datasets

We performed initial experiments on a dataset of Congressional debates that has annotations on the author level for partisanship, not ideology. While the two terms are highly correlated (e.g., a member of the Republican party likely agrees with conservative stances on most issues), they are not identical. For example, a moderate Republican might agree with the liberal position on increased gun control but take conservative positions on other issues. To avoid conflating partisanship and ideology we create a new dataset annotated for ideological bias on the sentence and phrase level. In this section we describe our initial dataset (Convote) and explain the procedure we followed for creating our new dataset (IBC).[1]

3.1 Convote

The Convote dataset (Thomas et al., 2006) consists of US Congressional floor debate transcripts from 2005 in which all speakers have been labeled with their political party (Democrat, Republican, or independent). We propagate party labels down from the speaker to all of their individual sentences and map from party label to ideology label (Democrat $\rightarrow$ liberal, Republican $\rightarrow$ conservative). This is an expedient choice; in future work we plan to make use of work in political science characterizing candidates' ideological positions empirically based on their behavior (Carroll et al., 2009).

While the Convote dataset has seen widespread use for document-level political classification, we are unaware of similar efforts at the sentence level.

3.1.1 Biased Sentence Selection

The strong correlation between US political parties and political ideologies (Democrats with liberal, Republicans with conservative) lends confidence that this dataset contains a rich mix of ideological

[1] Available at `http://cs.umd.edu/~miyyer/ibc`

statements. However, the raw Convote dataset contains a low percentage of sentences with explicit ideological bias.[2] We therefore use the features in Yano et al. (2010), which correlate with political bias, to select sentences to annotate that have a higher likelihood of containing bias. Their features come from the Linguistic Inquiry and Word Count lexicon (LIWC) (Pennebaker et al., 2001), as well as from lists of "sticky bigrams" (Brown et al., 1992) strongly associated with one party or another (e.g., "illegal aliens" implies conservative, "universal healthcare" implies liberal).

We first extract the subset of sentences that contains any words in the LIWC categories of Negative Emotion, Positive Emotion, Causation, Anger, and Kill verbs.[3] After computing a list of the top 100 sticky bigrams for each category, ranked by log-likelihood ratio, and selecting another subset from the original data that included only sentences containing at least one sticky bigram, we take the union of the two subsets. Finally, we balance the resulting dataset so that it contains an equal number of sentences from Democrats and Republicans, leaving us with a total of 7,816 sentences.

3.2 Ideological Books

In addition to Convote, we use the Ideological Books Corpus (IBC) developed by Gross et al. (2013). This is a collection of books and magazine articles written between 2008 and 2012 by authors with well-known political leanings. Each document in the IBC has been manually labeled with coarse-grained ideologies (right, left, and center) as well as fine-grained ideologies (e.g., religious-right, libertarian-right) by political science experts.

There are over a million sentences in the IBC, most of which have no noticeable political bias. Therefore we use the filtering procedure outlined in Section 3.1.1 to obtain a subset of 55,932 sentences. Compared to our final Convote dataset, an even larger percentage of the IBC sentences exhibit no noticeable political bias.[4] Because our goal is to distinguish between liberal and conservative bias, instead of the more general task of classifying sentences as "neutral" or "biased", we filter the dataset further using DUALIST (Settles, 2011), an active learning tool, to reduce the proportion of neutral sentences in our dataset. To train the DUALIST classifier, we manually assigned class labels of "neutral" or "biased" to 200 sentences, and selected typical partisan unigrams to represent the "biased" class. DUALIST labels 11,555 sentences as politically biased, 5,434 of which come from conservative authors and 6,121 of which come from liberal authors.

3.2.1 Annotating the IBC

For purposes of annotation, we define the task of political ideology detection as identifying, if possible, the political position of a given sentence's author, where position is either *liberal* or *conservative*.[5] We used the Crowdflower crowdsourcing platform (crowdflower.com), which has previously been used for subsentential sentiment annotation (Sayeed et al., 2012), to obtain human annotations of the filtered IBC dataset for political bias on both the sentence and phrase level. While members of the Crowdflower workforce are certainly not experts in political science, our simple task and the ubiquity of political bias allows us to acquire useful annotations.

Crowdflower Task First, we parse the filtered IBC sentences using the Stanford constituency parser (Socher et al., 2013a). Because of the expense of labeling every node in a sentence, we only label one path in each sentence. The process for selecting paths is as follows: first, if any paths contain one of the top-ten partisan unigrams,[6] we select the longest such path; otherwise, we select the path with the most open class constituencies (NP, VP, ADJP). The root node of a sentence is always included in a path.

Our task is shown in Figure 3. Open class constituencies are revealed to the worker incrementally, starting with the NP, VP, or ADJP furthest from the root and progressing up the tree. We choose this design to prevent workers from changing their lower-level phrase annotations after reading the full sentence.

[2] Many sentences in Convote are variations on "I think this is a good/bad bill", and there is also substantial parliamentary boilerplate language.

[3] While Kill verbs are not a category in LIWC, Yano et al. (2010) adopted it from Greene and Resnik (2009) and showed it to be a useful predictor of political bias. It includes words such as "slaughter" and "starve".

[4] This difference can be mainly attributed to a historical topics in the IBC (e.g., the Crusades, American Civil War). In Convote, every sentence is part of a debate about 2005 political policy.

[5] This is a simplification, as the ideological hierarchy in IBC makes clear.

[6] The words that the multinomial naïve Bayes classifier in DUALIST marked as highest probability given a polarity: market, abortion, economy, rich, liberal, tea, economic, taxes, gun, abortion

Filtering the Workforce To ensure our annotators have a basic understanding of US politics, we restrict workers to US IP addresses and require workers manually annotate one node from 60 different "gold " paths annotated by the authors. We select these nodes such that the associated phrase is either obviously biased or obviously neutral. Workers must correctly annotate at least six of eight gold paths before they are granted access to the full task. In addition, workers must maintain 75% accuracy on gold paths that randomly appear alongside normal paths. Gold paths dramatically improve the quality of our workforce: 60% of contributors passed the initial quiz (the 40% that failed were barred from working on the task), while only 10% of workers who passed the quiz were kicked out for mislabeling subsequent gold paths.

Annotation Results Workers receive the following instructions:

> Each task on this page contains a set of phrases from a single sentence. For each phrase, decide whether or not the author favors a political position to the left (*Liberal*) or right (*Conservative*) of center.
>
> - If the phrase is indicative of a position to the left of center, please choose *Liberal*.
> - If the phrase is indicative of a position to the right of center, please choose *Conservative*.
> - If you feel like the phrase indicates some position to the left or right of the political center, but you're not sure which direction, please mark *Not neutral, but I'm unsure of which direction*.
> - If the phrase is not indicative of a position to the left or right of center, please mark *Neutral*.

We had workers annotate 7,000 randomly selected paths from the filtered IBC dataset, with half of the paths coming from conservative authors and the other half from liberal authors, as annotated by Gross et al. (2013). Three workers annotated each path in the dataset, and we paid $0.03 per sentence. Since identifying political bias is a relatively difficult and subjective task, we include all sentences where at least two workers agree on a label for the root node in our final dataset, except when that label is "Not neutral, but I'm unsure of which direction". We only keep phrase-level annotations where at least two workers agree on the label: 70.4% of all annotated nodes fit this definition of agreement. All unannotated nodes receive the label of their closest annotated ancestor. Since the root of each sentence is always annotated, this strategy ensures that every node in the tree has a label. Our final balanced IBC dataset consists of 3,412 sentences (4,062 before balancing and removing neutral sentences) with a total of 13,640 annotated nodes. Of these sentences, 543 switch polarity (liberal $\rightarrow$ conservative or vice versa) on an annotated path.

Figure 3: Example political ideology annotation task showing incremental reveal of progressively longer phrases.

While we initially wanted to incorporate neutral labels into our model, we observed that lower-level phrases are almost always neutral while full sentences are much more likely to be biased (Figure 4). Due to this discrepancy, the objective function in Eq. (5) was minimized by making neutral predictions for almost every node in the dataset.

4 Experiments

In this section we describe our experimental framework. We discuss strong baselines that use lexical and syntactic information (including framing-specific features from previous work) as well as multiple RNN configurations. Each of these models have the same task: to predict sentence-level ideology labels for sentences in a test set. To account for label imbalance, we subsample the data so that there are an equal number of labels and report accuracy over this balanced dataset.

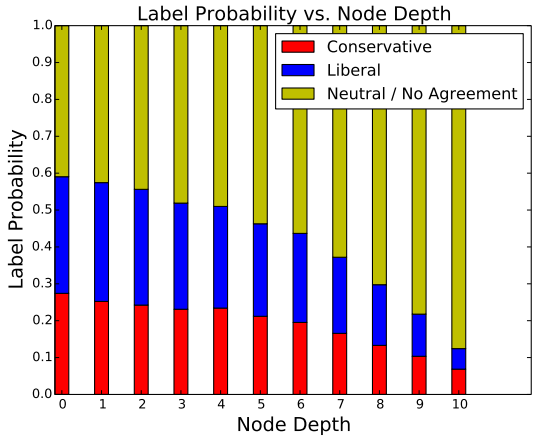

Figure 4: Proportion of liberal, conservative, and neutral annotations with respect to node depth (distance from root). As we get farther from the root of the tree, nodes are more likely to be neutral.

4.1 Baselines

- The **RANDOM** baseline chooses a label at random from {*liberal*, *conservative*}.
- **LR1**, our most basic logistic regression baseline, uses only bag of words (*BoW*) features.
- **LR2** uses only *BoW* features. However, **LR2** also includes phrase-level annotations as separate training instances.[7]
- **LR3** uses *BoW* features as well as syntactic pseudo-word features from Greene & Resnik (2009). These features from dependency relations specify properties of verbs (e.g., transitivity or nominalization).[8]
- **LR-(W2V)** is a logistic regression model trained on the average of the pretrained word embeddings for each sentence (Section 2.2).

The **LR-(W2V)** baseline allows us to compare against a strong lexical representation that encodes syntactic and semantic information without the RNN tree structure. (**LR1**, **LR2**) offer a comparison to simple bag of words models, while the **LR3** baseline contrasts traditional syntactic features with those learned by RNN models.

4.2 RNN Models

For RNN models, we generate a feature vector for every node in the tree. Equation 1 allows us to percolate the representations to the root of the tree. We generate the final instance representation by concatenating the root vector and the average of all other vectors (Socher et al., 2011b). We train an L_2-regularized logistic regression model over these concatenated vectors to obtain final accuracy numbers on the sentence level.

To analyze the effects of initialization and phrase-level annotations, we report results for three different RNN settings. All three models were implemented as described in Section 2 with the nonlinearity f set to the normalized $\tanh$ function,

$$f(v) = \frac{\tanh(v)}{\|\tanh(v)\|}. \quad (7)$$

We perform 10-fold cross-validation on the training data to find the best RNN hyperparameters.[9]

We report results for RNN models with the following configurations:

- **RNN1** initializes all parameters randomly and uses only sentence-level labels for training.
- **RNN1-(W2V)** uses the word2vec initialization described in Section 2.2 but is also trained on only sentence-level labels.
- **RNN2-(W2V)** is initialized using word2vec embeddings and also includes annotated phrase labels in its training. For this model, we also introduce a hyperparameter β that weights the error at annotated nodes $(1-\beta)$ higher than the error at unannotated nodes (β); since we have more confidence in the annotated labels, we want them to contribute more towards the objective function.

Model	Convote	IBC
RANDOM	50%	50%
LR1	64.7%	62.1%
LR2	–	61.9%
LR3	66.9%	62.6%
LR-(W2V)	66.6%	63.7%
RNN1	69.4%	66.2%
RNN1-(W2V)	**70.2%**	67.1%
RNN2-(W2V)	–	**69.3%**

Table 1: Sentence-level bias detection accuracy. The RNN framework, adding phrase-level data, and initializing with word2vec all improve performance over logistic regression baselines. The **LR2** and **RNN2-(W2V)** models were not trained on Convote since it lacks phrase annotations.

[7] The Convote dataset was not annotated on the phrase level, so we only provide a result for the IBC dataset.

[8] We do not include phrase-level annotations in the **LR3** feature set because the pseudo-word features can only be computed from full sentence parses.

[9] [λ_{W_e} =1e-6, λ_W =1e-4, $\lambda_{W_{cat}}$ =1e-3, $\beta = 0.3$]

For all RNN models, we set the word vector dimension d to 300 to facilitate direct comparison against the **LR-(W2V)** baseline.[10]

5 Where Compositionality Helps Detect Ideological Bias

In this section, we examine the RNN models to see why they improve over our baselines. We also give examples of sentences that are correctly classified by our best RNN model but incorrectly classified by all of the baselines. Finally, we investigate sentence constructions that our model cannot handle and offer possible explanations for these errors.

Experimental Results Table 1 shows the RNN models outperforming the bag-of-words baselines as well as the word2vec baseline on both datasets. The increased accuracy suggests that the trained RNNs are capable of detecting bias polarity switches at higher levels in parse trees. While phrase-level annotations do not improve baseline performance, the RNN model significantly benefits from these annotations because the phrases are themselves derived from nodes in the network structure. In particular, the phrase annotations allow our best model to detect bias accurately in complex sentences that the baseline models cannot handle.

Initializing the RNN W_e matrix with word2vec embeddings improves accuracy over randomly initialization by 1%. This is similar to improvements from pretrained vectors from neural language models (Socher et al., 2011b).

We obtain better results on Convote than on IBC with both bag-of-words and RNN models. This result was unexpected since the Convote labels are noisier than the annotated IBC labels; however, there are three possible explanations for the discrepancy. First, Convote has twice as many sentences as IBC, and the extra training data might help the model more than IBC's better-quality labels. Second, since the sentences in Convote were originally spoken, they are almost half as short (21.3 words per sentence) as those in the IBC (42.2 words per sentence). Finally, some information is lost at every propagation step, so RNNs are able to model the shorter sentences in Convote more effectively than the longer IBC sentences.

Qualitative Analysis As in previous work (Socher et al., 2011b), we visualize the learned vector space by listing the most probable n-grams for each political affiliation in Table 2. As expected, conservatives emphasize values such as freedom and religion while disparaging excess government spending and their liberal opposition. Meanwhile, liberals inveigh against the gap between the rich and the poor while expressing concern for minority groups and the working class.

Our best model is able to accurately model the compositional effects of bias in sentences with complex syntactic structures. The first three sentences in Figure 5 were correctly classified by our best model (**RNN2-(W2V)**) and incorrectly classified by all of the baselines. Figures 5A and C show traditional conservative phrases, "free market ideology" and "huge amounts of taxpayer money", that switch polarities higher up in the tree when combined with phrases such as "made worse by" and "saved by". Figure 5B shows an example of a bias polarity switch in the opposite direction: the sentence negatively portrays supporters of nationalized health care, which our model picks up on.

Our model often makes errors when polarity switches occur at nodes that are high up in the tree. In Figure 5D, "be used as an instrument to achieve charitable or social ends" reflects a liberal ideology, which the model predicts correctly. However, our model is unable to detect the polarity switch when this phrase is negated with "should not". Since many different issues are discussed in the IBC, it is likely that our dataset has too few examples of some of these issues for the model to adequately learn the appropriate ideological positions, and more training data would resolve many of these errors.

6 Related Work

A growing NLP subfield detects private states such as opinions, sentiment, and beliefs (Wilson et al., 2005; Pang and Lee, 2008) from text. In general, work in this category tends to combine traditional surface lexical modeling (e.g., bag-of-words) with hand-designed syntactic features or lexicons. Here we review the most salient literature related to the present paper.

6.1 Automatic Ideology Detection

Most previous work on ideology detection ignores the syntactic structure of the language in use in favor of familiar bag-of-words representations for

[10] Using smaller vector sizes ($d \in \{50, 100\}$, as in previous work) does not significantly change accuracy.

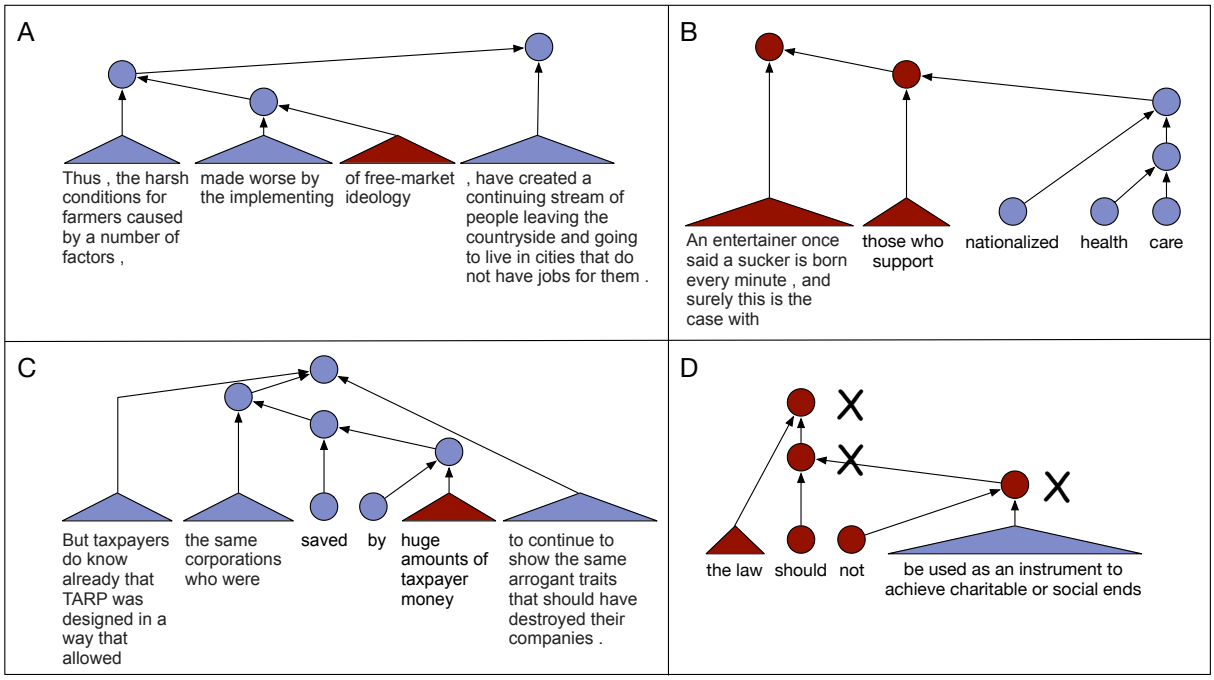

Figure 5: Predictions by **RNN2-(W2V)** on four sentences from the IBC. Node color is the true label (red for conservative, blue for liberal), and an "X" next to a node means the model's prediction was wrong. In *A* and *C*, the model accurately detects conservative-to-liberal polarity switches, while in *B* it correctly predicts the liberal-to-conservative switch. In *D*, negation confuses our model.

the sake of simplicity. For example, Gentzkow and Shapiro (2010) derive a "slant index" to rate the ideological leaning of newspapers. A newspaper's slant index is governed by the frequency of use of partisan collocations of 2-3 tokens. Similarly, authors have relied on simple models of language when leveraging inferred ideological positions. E.g., Gerrish and Blei (2011) predict the voting patterns of Congress members based on bag-of-words representations of bills and inferred political leanings of those members.

Recently, Sim et al. (2013) have proposed a model to infer mixtures of ideological positions in documents, applied to understanding the evolution of ideological rhetoric used by political candidates during the campaign cycle. They use an HMM-based model, defining the states as a set of fine-grained political ideologies, and rely on a closed set of lexical bigram features associated with each ideology, inferred from a manually labeled ideological books corpus. Although it takes elements of discourse structure into account (capturing the"burstiness" of ideological terminology usage), their model explicitly ignores intrasentential contextual influences of the kind seen in Figure 1. Other approaches on the document level use topic models to analyze bias in news articles, blogs, and political speeches (Ahmed and Xing, 2010; Lin et al., 2008; Nguyen et al., 2013).

6.2 Subjectivity Detection

Detecting subjective language, which conveys opinion or speculation, is a related NLP problem. While sentences lacking subjective language may contain ideological bias (e.g., the topic of the sentence), highly-opinionated sentences likely have obvious ideological leanings. In addition, sentiment and subjectivity analysis offers methodological approaches that can be applied to automatic bias detection.

Wiebe et al. (2004) show that low-frequency words and some collocations are a good indicators of subjectivity. More recently, Recasens et al. (2013) detect biased words in sentences using indicator features for bias cues such as hedges and factive verbs in addition to standard bag-of-words and part-of-speech features. They show that this type of linguistic information dramatically improves performance over several standard baselines.

Greene and Resnik (2009) also emphasize the connection between syntactic and semantic relationships in their work on "implicit sentiment",

n	Most conservative n-grams	Most liberal n-grams
1	Salt, Mexico, housework, speculated, consensus, lawyer, pharmaceuticals, ruthless, deadly, Clinton, redistribution	rich, antipsychotic, malaria, biodiversity, richest, gene, pesticides, desertification, Net, wealthiest, labor, fertilizer, nuclear, HIV
3	prize individual liberty, original liberal idiots, stock market crash, God gives freedom, federal government interference, federal oppression nullification, respect individual liberty, Tea Party patriots, radical Sunni Islamists, Obama stimulus programs	rich and poor,"corporate greed", super rich pay, carrying the rich, corporate interest groups, young women workers, the very rich, for the rich, by the rich, soaking the rich, getting rich often, great and rich, the working poor, corporate income tax, the poor migrants
5	spending on popular government programs, bailouts and unfunded government promises, North America from external threats, government regulations place on businesses, strong Church of Christ convictions, radical Islamism and other threats	the rich are really rich, effective forms of worker participation, the pensions of the poor, tax cuts for the rich, the ecological services of biodiversity, poor children and pregnant women, vacation time for overtime pay
7	government intervention helped make the Depression Great, by God in His image and likeness, producing wealth instead of stunting capital creation, the traditional American values of limited government, trillions of dollars to overseas oil producers, its troubled assets to federal sugar daddies, Obama and his party as racialist fanatics	African Americans and other disproportionately poor groups; the growing gap between rich and poor; the Bush tax cuts for the rich; public outrage at corporate and societal greed; sexually transmitted diseases , most notably AIDS; organize unions or fight for better conditions, the biggest hope for health care reform

Table 2: Highest probability n-grams for conservative and liberal ideologies, as predicted by the **RNN2-(W2V)** model.

which refers to sentiment carried by sentence structure and not word choice. They use syntactic dependency relation features combined with lexical information to achieve then state-of-the-art performance on standard sentiment analysis datasets. However, these syntactic features are only computed for a thresholded list of domain-specific verbs. This work extends their insight of modeling sentiment as an interaction between syntax and semantics to ideological bias.

Future Work There are a few obvious directions in which this work can be expanded. First, we can consider more nuanced political ideologies beyond *liberal* and *conservative*. We show that it is possible to detect ideological bias given this binary problem; however, a finer-grained study that also includes neutral annotations may reveal more subtle distinctions between ideologies. While acquiring data with obscure political biases from the IBC or Convote is unfeasible, we can apply a similar analysis to social media (e.g., Twitter or Facebook updates) to discover how many different ideologies propagate in these networks.

Another direction is to implement more sophisticated RNN models (along with more training data) for bias detection. We attempted to apply syntactically-untied RNNs (Socher et al., 2013a) to our data with the idea that associating separate matrices for phrasal categories would improve representations at high-level nodes. While there were too many parameters for this model to work well here, other variations might prove successful, especially with more data. Finally, combining sentence-level and document-level models might improve bias detection at both levels.

7 Conclusion

In this paper we apply recursive neural networks to political ideology detection, a problem where previous work relies heavily on bag-of-words models and hand-designed lexica. We show that our approach detects bias more accurately than existing methods on two different datasets. In addition, we describe an approach to crowdsourcing ideological bias annotations. We use this approach to create a new dataset from the IBC, which is labeled at both the sentence and phrase level.

Acknowledgments

We thank the anonymous reviewers, Hal Daumé, Yuening Hu, Yasuhiro Takayama, and Jyothi Vinjumur for their insightful comments. We also want to thank Justin Gross for providing the IBC and Asad Sayeed for help with the Crowdflower task design, as well as Richard Socher and Karl Moritz Hermann for assisting us with our model implementations. This work was supported by NSF Grant CCF-1018625. Boyd-Graber is also supported by NSF Grant IIS-1320538. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the view of the sponsor.

References

Amr Ahmed and Eric P Xing. 2010. Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective. In *EMNLP*.

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based n-gram models of natural language. *Comp. Ling.*, 18(4):467–479.

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 and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *ICML*.

Frank E Dardis, Frank R Baumgartner, Amber E Boydstun, 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 *ICML*.

Christoph Goller and Andreas Kuchler. 1996. Learning task-dependent distributed representations by backpropagation through structure. In *Neural Networks, 1996., IEEE International Conference on*, volume 1.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *NAACL*.

Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.

Justin Gross, Brice Acree, Yanchuan Sim, and Noah A Smith. 2013. Testing the etch-a-sketch hypothesis: A computational analysis of mitt romney's ideological makeover during the 2012 primary vs. general elections. In *APSA 2013 Annual Meeting Paper*.

Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. 2013. Simple customization of recursive neural networks for semantic relation classification. In *EMNLP*.

Karl Moritz Hermann and Phil Blunsom. 2013. The Role of Syntax in Vector Space Models of Compositional Semantics. In *ACL*.

George Lakoff. 2002. *Moral Politics: How Liberals and Conservatives Think, Second Edition*. University of Chicago Press.

Wei-Hao Lin, Eric Xing, and Alexander Hauptmann. 2008. A joint topic and perspective model for ideological discourse. In *Machine Learning and Knowledge Discovery in Databases*, pages 17–32. Springer.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik. 2013. Lexical and hierarchical topic regression. In *NIPS*, pages 1106–1114.

David Niven. 2003. Objective evidence on media bias: Newspaper coverage of congressional party switchers. *Journalism & Mass Communication Quarterly*, 80(2):311–326.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2).

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. *Linguistic inquiry and word count [computer software]*. Mahwah, NJ: Erlbaum Publishers.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language.

Asad B Sayeed, Jordan Boyd-Graber, Bryan Rusk, and Amy Weinberg. 2012. Grammatical structures for word-level sentiment detection. In *NAACL*.

Burr Settles. 2011. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *EMNLP*.

Yanchuan Sim, Brice Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *EMNLP*.

Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011a. Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection. In *NIPS*.

Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011b. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In *EMNLP*.

Richard Socher, John Bauer, Christopher D. Manning, and Andrew Y. Ng. 2013a. Parsing With Compositional Vector Grammars. In *ACL*.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013b. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *EMNLP*.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Comp. Ling.*, 30(3):277–308.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP*.

Tae Yano, Philip Resnik, and Noah A Smith. 2010. Shedding (a thousand points of) light on biased language. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 152–158.