Beyond Binary Labels: Political Ideology Prediction of Twitter Users

Daniel Preot˘iuc-Pietro
Positive Psychology Center
University of Pennsylvania
danielpr@sas.upenn.edu

Ye Liu∗
School of Computing
National University of Singapore
liuye@comp.nus.edu.sg

Daniel J. Hopkins
Political Science Department
University of Pennsylvania
danhop@sas.upenn.edu

Lyle Ungar
Computing & Information Science
University of Pennsylvania
ungar@cis.upenn.edu

Abstract

Automatic political preference prediction from social media posts has to date proven successful only in distinguishing between publicly declared liberals and conservatives in the US. This study examines users’ political ideology using a seven-point scale which enables us to identify politically moderate and neutral users – groups which are of particular interest to political scientists and pollsters. Using a novel data set with political ideology labels self-reported through surveys, our goal is two-fold: a) to characterize the political groups of users through language use on Twitter; b) to build a fine-grained model that predicts political ideology of unseen users. Our results identify differences in both political leaning and engagement and the extent to which each group tweets using political keywords. Finally, we demonstrate how to improve ideology prediction accuracy by exploiting the relationships between the user groups.

1 Introduction

Social media is used by people to share their opinions and views. Unsurprisingly, an important part of the population shares opinions and news related to politics or causes they support, thus offering strong cues about their political preferences and ideologies. In addition, political membership is also predictable purely from one’s interests or demographics — it is much more likely for a religious person to be conservative or for a younger person to lean liberal (Ellis and Stimson, 2012).

User trait prediction from text is based on the assumption that language use reflects a user’s demographics, psychological states or preferences. Applications include prediction of age (Rao et al., 2010; Flekova et al., 2016b), gender (Burger et al., 2011; Sap et al., 2014), personality (Schwartz et al., 2013; Preot˘iuc-Pietro et al., 2016), socioeconomic status (Preot˘iuc-Pietro et al., 2015a;b; Liu et al., 2016c), popularity (Lampos et al., 2014) or location (Cheng et al., 2010).

Research on predicting political orientation has focused on methodological improvements (Penacchiotti and Popescu, 2011) and used data sets with publicly stated dichotomous political orientation labels due to their easy accessibility (Sylwester and Purver, 2015). However, these data sets are not representative samples of the entire population (Cohen and Ruths, 2013) and do not accurately reflect the variety of political attitudes and engagement (Kam et al., 2007).

For example, we expect users who state their political affiliation in their profile description, tweet with partisan hashtags or appear in public party lists to use social media as a means of popularizing and supporting their political beliefs (BarberASa, 2015). Many users may choose not to publicly post about their political preference for various social goals or perhaps this preference may not be strong or representative enough to be disclosed online. Dichotomous political preference also ignores users who do not have a political ideology. All of these types of users are very important for researchers aiming to understand group preferences, traits or moral values (Lewis and Reiley, 2014; Hersh, 2015).

The most common political ideology spectrum in the US is the conservative – liberal (Ellis and Stimson, 2012). We collect a novel data set of Twitter users mapped to this seven-point spectrum which allows us to:

∗ Work carried out during a research visit at the University of Pennsylvania
1. Uncover the differences in language use between ideological groups;
2. Develop a user-level political ideology prediction algorithm that classifies all levels of engagement and leverages the structure in the political ideology spectrum.

First, using a broad range of language features including unigrams, word clusters and emotions, we study the linguistic differences between the two ideologically extreme groups, the two ideologically moderate groups and between both extremes and moderates in order to provide insight into the content they post on Twitter. In addition, we examine the extent to which the ideological groups in our data set post about politics and compare it to a data set obtained similarly to previous work.

In prediction experiments, we show how accurately we can distinguish between opposing ideological groups in various scenarios and that previous binary political orientation prediction has been oversimplified. Then, we measure the extent to which we can predict the two dimensions of political leaning and engagement. Finally, we build an ideology classifier in a multi-task learning setup that leverages the relationships between groups.\footnote{Data is available at http://www.preotiuc.ro}

### 2 Related Work

Automatically inferring user traits from their online footprints is a prolific topic of research, enabled by the increasing availability of user-generated data and advances in machine learning. Beyond its research oriented goals, user profiling has important industry applications in online marketing, personalization or large-scale audience profiling. To this end, researchers have used a wide range of types of online footprints, including video (Subramanian et al., 2013), audio (Alam and Riccardi, 2014), text (Preotiuc-Pietro et al., 2015a), profile images (Liu et al., 2016a), social data (Van Der Heide et al., 2012; Hall et al., 2014), social networks (Perozzi and Skiena, 2015; Rout et al., 2013), payment data (Wang et al., 2016) and endorsements (Kosinski et al., 2013).

Political orientation prediction has been studied in two related, albeit crucially different scenarios, as also identified in (Zafar et al., 2016). First, researchers aimed to identify and quantify orientation of words (Monroe et al., 2008), hashtags (Weber et al., 2013) or documents (Iyyer et al., 2014), or to detect bias (Yano et al., 2010) or impartiality (Zafar et al., 2016) at a document level.

Our study belongs to the second category, where political orientation is inferred at a user-level. All previous studies study labeling US conservatives vs. liberals using either text (Rao et al., 2010), social network connections (Zamal et al., 2012), platform-specific features (Conover et al., 2011) or a combination of these (Pennacchiotti and Popescu, 2011; Volkova et al., 2014), with very high reported accuracies of up to 94.9% (Conover et al., 2011).

However, all previous work on predicting user-level political preferences are limited to a binary prediction between liberal/democrat and conservative/republican, disregarding any nuances in political ideology. In addition, as the focus of the studies is more on the methodological or interpretation aspects of the problem, another downside is that the user labels were obtained in simple, albeit biased ways. These include users who explicitly state their political orientation on user lists of party supporters (Zamal et al., 2012; Pennacchiotti and Popescu, 2011), supporting partisan causes (Rao et al., 2010), by following political figures (Volkova et al., 2014) or party accounts (Sylwester and Purver, 2015) or that retweet partisan hashtags (Conover et al., 2011). As also identified in (Cohen and Ruths, 2013) and further confirmed later in this study, these data sets are biased: most people do not clearly state their political preference online – fewer than 5% according to Priante et al. (2016) – and those that state their preference are very likely to be political activists. Cohen and Ruths (2013) demonstrated that predictive accuracy of classifiers is significantly lower when confronted with users that do not explicitly mention their political orientation. Despite this, their study is limited because in their hardest classification task, they use crowdsourced political orientation labels, which may not correspond to reality and suffer from biases (Flekova et al., 2016a; Carpenter et al., 2016). Further, they still only look at predicting binary political orientation. To date, no other research on this topic has taken into account these findings.

### 3 Data Set

The main data set used in this study consists of 3,938 users recruited through the Qualtrics platform ($D_1$). Each participant was compensated...
with 3 USD for 15 minutes of their time. All participants first answered the same demographic questions (including political ideology), then were directed to one of four sets of psychological questionnaires unrelated to the political ideology question. They were asked to self-report their political ideology on a seven point scale: Very conservative (1), Conservative (2), Moderately conservative (3), Moderate (4), Moderately liberal (5), Liberal (6), Very liberal (7). In addition, participants had the option of choosing Apathetic and Other, which have ambiguous fits on the conservative – liberal spectrum and were removed from our analysis (399 users). We also asked participants to self-report their gender (2322 female, 1205 male, 12 other) and age. Participants were all from the US in order to limit the impact of cultural and political factors. The political ideology distribution in our sample is presented in Figure 1.

We asked users their Twitter handle and downloaded their most recent 3,200 tweets, leading to a total of 4,833,133 tweets. Before adding users to our 3,938 user data set, we performed the following checks to ensure that the Twitter handle was the user’s own: 1) after compensation, users were if they were truthful in reporting their handle and if not, we removed their data from analysis; 2) we manually examined all handles marked as verified by Twitter or that had over 2000 followers and eliminated them if they were celebrities or corporate/news accounts, as these were unlikely the users who participated in the survey. This study received approval from the Institutional Review Board (IRB) of the University of Pennsylvania.

In addition, to facilitate comparison to previous work, we also use a data set of 13,651 users with overt political orientation ($D_2$). We selected popular political figures unambiguously associated with US liberal politics (@SenSanders, @JoeBiden, @CoryBooker, @JohnKerry) or US conservative politics (@marcorubio, @tedcruz, @RandPaul, @RealBenCarson). Liberals in our set ($N_l = 7417$) had to follow on Twitter all of the liberal political figures and none of the conservative figures. Likewise, conservative users ($N_c = 6234$) had to follow all of the conservative figures and no liberal figures. We downloaded up to 3,200 of each user’s most recent tweets, leading to a total of 25,493,407 tweets. All tweets were downloaded around 10 August 2016.

4 Features

In our analysis, we use a broad range of linguistic features described below.

Unigrams We use the bag-of-words representation to reduce each user’s posting history to a normalised frequency distribution over the vocabulary consisting of all words used by at least 10% of the users (6,060 words).

LIWC Traditional psychological studies use a dictionary-based approach to representing text. The most popular method is based on Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001), and automatically counts word frequencies for 64 different categories manually constructed based on psychological theory. These include different parts-of-speech, topical categories and emotions. Each user is thereby represented as a frequency distribution over these categories.

Word2Vec Topics An alternative to LIWC is to use automatically generated word clusters i.e., groups of words that are semantically and/or syntactically similar. The clusters help reducing the feature space and provides additional interpretability.

To create these groups of words, we use an automatic method that leverages word co-occurrence patterns in large corpora by making use of the distributional hypothesis: similar words tend to co-occur in similar contexts (Harris, 1954). Based on co-occurrence statistics, each word is represented as a low dimensional vector of numbers with words closer in this space being more similar (Deerwester et al., 1990). We use the method from (Preotiuc-Pietro et al., 2015a) to compute topics using word2vec similarity (Mikolov et al., 2013a,b) and spectral clustering (Shi and Malik, 2000; von Luxburg, 2007) of different sizes (from 30 to 2000). We have tried other alternatives to building clusters: using other word similarities to
generate clusters – such as NPMI (Lampos et al., 2014) or GloVe (Pennington et al., 2014) as proposed in (Preotiuc-Pietro et al., 2015a) – or using standard topic modelling approaches to create soft clusters of words e.g., Latent Dirichlet Allocation (Blei et al., 2003). For brevity, we present experiments with the best performing feature set containing 500 Word2Vec clusters. We aggregate all the words posted in a user’s tweets and represent each user as a distribution of the fraction of words belonging to each cluster.

Sentiment & Emotions We hypothesise that different political ideologies differ in the type and amount of emotions the users express through their posts. The most studied model of discrete emotions is the Ekman model (Ekman, 1992; Strapparava and Mihalcea, 2008; Strapparava et al., 2004) which posits the existence of six basic emotions: anger, disgust, fear, joy, sadness and surprise. We automatically quantify these emotions from our Twitter data set using a publicly available crowd-sourcing derived lexicon of words associated with any of the six emotions, as well as general positive and negative sentiment (Mohammad and Turney, 2010, 2013). Using these lexicons, we assign a predicted emotion to each message and then average across all users’ posts to obtain user level emotion expression scores.

Political Terms In order to select unigrams pertaining to politics, we assigned our most frequent 12,000 unigrams in our data set to three categories:

- **Political words**: mentions of political terms (234);
- **Political NEs**: mentions of politician proper names out of the political terms (39);
- **Media NEs**: mentions of political media sources and pundits out of the political terms (20).

This coding was initially performed by a research assistant studying political science with good knowledge of US politics and were further filtered and checked by one of the authors.

5 Analysis

First, we explore the relationships between language use and political ideological groups within each feature set and pairs of opposing user groups. To illustrate differences between ideological groups we compare the two political extremes (Very Conservative – Very Liberal) and the political moderates (Moderate Conservative – Moderate Liberal). We further compare outright moderates with a group combining the two political extremes to study if we can uncover differences in political engagement and extremity, regardless of the conservative–liberal leaning.

We use univariate partial linear correlations with age and gender as co-variates to factor out the influence of basic demographics. For example, in $D_1$, users who reported themselves as very conservative are older and more likely males ($\mu_{age} = 35.1, \text{pct\_male} = 44\%$) than the data average ($\mu_{age} = 31.2, \text{pct\_male} = 35\%$). Additionally, prior to combining the two ideologically extreme groups, we sub-sampled the larger class (Very Liberal) to match the smaller class (Very Conservative) in age and gender. In the later prediction experiments, we do not perform matching, as this represents useful signal for classification (Ellis and Stimson, 2012). Results with unigrams are presented in Figure 2 and with the other features in Table 1. These are selected using standard statistical significance tests.

5.1 Very Conservatives vs. Very Liberals

The comparison between the extreme categories reveals the largest number of significant differences. The unigrams and Word2Vec clusters specific to conservatives are dominated by religion specific terms (‘praying’, ‘god’, W2V-485, W2V-018, W2V-099, L-RELIG), confirming a well-documented relationship (Gelman, 2009) and words describing family relationships (‘uncle’, ‘son’, L-FAMILY), another conservative value (Lakoff, 1997). The emphasis on religious terms among conservatives is consistent with the claim that many Americans associate ‘conservative’ with ‘religious’ (Ellis and Stimson, 2012). Extreme liberals show a tendency to use more adjectives (W2V-075, W2V-110), adverbs (L-ADVERB), conjunctions (L-CONJ) and comparisons (L-COMPARE) which indicate more nuanced and complex posts. Extreme conservatives post tweets higher in all positive emotions than liberals (L-POSEMO, Emot-Joy, Emot-Positive), confirming a previously hypothesised relationship (Napier and Jost, 2008). However, extreme liberals are not associated with posting negative emotions either, only using words that reflect more anxiety (L-ANX), which is related to neuroticism in which the liberals are higher (Gerber et al., 2010).

Political term analysis reveals the partisan terms
employed by both sides. For example, conservatives retweet or mention politicians such as Donald Trump or Ted Cruz, while liberals mention Barack Obama. Extreme conservatives also reference known partisan conservative media sources (@foxnews, @yahoonews) and hashtags (#pajnet,
#tcot), while extreme liberals focus on issues (‘gay’, ‘racism’, ‘feminism’, ‘transgender’). This perhaps reflects the desire for conservatives on Twitter to identify like-minded individuals, as extreme conservatives are a minority on the platform. Liberals, by contrast, use the platform to discuss and popularize their causes.

5.2 Moderate Conservatives vs. Moderate Liberals

Comparing the two sides of moderate users reveals a slightly more nuanced view of the two ideologies. While moderate conservatives still make heavy use of religious terms and express positive emotions (Emot-Joy, L-DRIVES), they also use affiliative language (L-AFFILIATION) and plural pronouns (L-WE). Moderate liberals are identified by very different features compared to their more extreme counterparts. Most striking is the use of swear and sex words (L-SEXUAL, L-ANGER, W2V-316), also highlighted by Sylvester and Purver (2015). Two word clusters relating to British culture (W2V-458) and art (W2V-373) reflect that liberals are more inclined towards arts (Dollinger, 2007). Statistically significant political terms are very few compared to the previous comparison, probably due to their lower overall usage, which we further investigate later.

5.3 Moderates vs. Extremists

Our final comparison looks at outright moderates compared to the two extreme groups combined, as we hypothesise the existence of a difference in overall political engagement. Moderates are not characterized by many features besides a topic of casual words (W2V-098), indicating the heterogeneity of this group of users. However, regardless of their orientation, the ideological extremists stand out from moderates. They use words and word clusters related to political actors (W2V-309), issues (W2V-237) and laws (W2V-296, W2V-288). LIWC analysis uncovers differences in article use (L-ARTICLE) or power words (L-POWER) specific of political tweets. The overall sentiment of these users is negative (Emot-Fear, Emot-Disgust, Emot-Sadness, L-DEATH) compared to moderates. This reveals – combined with the finding from the first comparison – that while extreme conservatives are overall more positive than liberals, both groups share negative expression. Political terms are almost all significantly correlated with the extreme ideological groups, confirming the existence of a difference in political engagement which we study in detail next.

5.4 Political Terms

Figure 3 presents the use of the three types of political terms across the 7 ideological groups in $D_1$ and the two political groups from $D_2$. We notice the following:

- $D_2$ has a huge skew towards political words, with an average of more than three times more political terms across all three categories than our extreme classes from $D_1$;
- Within the groups in $D_1$, we observe an almost perfectly symmetrical U-shape across all three types of political terms, confirming our hypothesis about political engagement;
- The difference between 1–2/6–7 is larger than 2–3/5–6. The extreme liberals and conservatives are disproportionately political, and have the potential to give Twitter’s political discussions an unrepresentative, extremist hue (Fiorina, 1999). It is also possible, however, that characterizing one as an extreme liberal or conservative indicates as much about her level of political engagement as it does about her placement on a left-right scale (Converse, 1964; Broockman, 2016).

6 Prediction

In this section we build predictive models of political ideology and compare them to data sets obtained using previous work.
6.1 Cross-Group Prediction

First, we experiment with classifying between conservatives and liberals across various levels of political engagement in \( D_1 \) and between the two polarized groups in \( D_2 \). We use logistic regression classification to compare three setups in Table 2 with results measured with ROC AUC as the classes are slightly imbalanced:

- 10-fold cross-validation where training is performed on the same task as the testing (principal diagonal);
- A train–test setup where training is performed on one task (presented in rows) and testing is performed on another (presented in columns);
- A domain adaptation setup (results in brackets) where on each of the 10 folds, the 9 training folds (presented in rows) are supplemented with all the data from a different task (presented in columns) using the EasyAdapt algorithm (Daumé III, 2007) as a proof on concept on the effects of using additional distantly supervised data. Data pooling lead to worse results than EasyAdapt.

Each of the three tasks from \( D_1 \) have a similar number of training samples, hence we do not expect that data set size has any effects in comparing the results across tasks.

The results with both sets of features show that:

- Prediction performance is much higher for \( D_2 \) than for \( D_1 \), with the more extreme groups in \( D_1 \) being easier to predict than the moderate groups. This confirms that the very high accuracies reported by previous research are an artifact of user label collection and that on regular users, the expected accuracy is much lower (Cohen and Ruths, 2013). We further show that, as the level of political engagement decreases, the classification problem becomes even harder;
- The model trained on \( D_2 \) and Word2Vec word clusters performs significantly worse on \( D_1 \) tasks even if the training data is over 10 times larger. When using political words, the \( D_2 \) trained classifier performs relatively well on all tasks from \( D_1 \);
- Overall, using political words as features performs better than Word2Vec clusters in the binary classification tasks;
- Domain adaptation helps in the majority of cases, leading to improvements of up to .03 in AUC (predicting 2v6 supplemented with 3v5 data).

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1v7</td>
<td>2v6</td>
</tr>
<tr>
<td>.785</td>
<td>.639 (.681)</td>
</tr>
<tr>
<td>.729 (.789)</td>
<td>.662</td>
</tr>
<tr>
<td>.618 (.778)</td>
<td>.617 (.690)</td>
</tr>
<tr>
<td>.708 (.764)</td>
<td>.627 (.644)</td>
</tr>
</tbody>
</table>

(a) Word2Vec 500

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1v7</td>
<td>2v6</td>
</tr>
<tr>
<td>.785</td>
<td>.551 (.679)</td>
</tr>
<tr>
<td>.739 (.773)</td>
<td>.679</td>
</tr>
<tr>
<td>.727 (.766)</td>
<td>.636 (.670)</td>
</tr>
<tr>
<td>.766 (.789)</td>
<td>.677 (.683)</td>
</tr>
</tbody>
</table>

(b) Political Terms

<table>
<thead>
<tr>
<th>1v7</th>
<th>2v6</th>
<th>3v5</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>.785</td>
<td>.639 (.681)</td>
<td>.575 (.598)</td>
<td>.705 (.887)</td>
</tr>
<tr>
<td>.729 (.789)</td>
<td>.662</td>
<td>.574 (.856)</td>
<td>.663 (.889)</td>
</tr>
<tr>
<td>.618 (.778)</td>
<td>.617 (.690)</td>
<td>.581</td>
<td>.684 (.887)</td>
</tr>
<tr>
<td>.708 (.764)</td>
<td>.627 (.644)</td>
<td>.571 (.574)</td>
<td>.891</td>
</tr>
</tbody>
</table>

Table 2: Prediction results of the logistic regression classification in ROC AUC when discriminating between two political groups across different levels of engagement and both data sets. The binary classifier from data set \( D_2 \) is represented by \( D_2 \), the rest of the categories are from data set \( D_1 \). Results on the principal diagonal represent 10-fold cross-validation results (training in-domain). Results off-diagonal represent training the classifier from the column and testing on the problem indicated in the row (training out-of-domain). Numbers in brackets indicate performance when the training data was added in the 10-fold cross-validation setup using the EasyAdapt algorithm (domain adaptation). Best results without domain adaptation are in bold, while the best results with domain adaptation are in italics.

6.2 Political Leaning and Engagement Prediction

Political leaning (Conservative – Liberal, excluding the Moderate group) can be considered an ordinal variable and the prediction problem framed as one of regression. In addition to the political leaning prediction, based on analysis and previous prediction results, we hypothesize the existence of a separate dimension of political engagement regardless of the partisan side. Thus, we merge users from classes 3–5, 2–6, 1–7 and create a variable with four values, where the lowest value is represented by moderate users (4) and the highest value is represented by either very conservative (1) or very liberal (7) users.

We use a linear regression algorithm with an Elastic Net regularizer (Zou and Hastie, 2005) as implemented in ScikitLearn (Pedregosa et al., 2011). To evaluate our results, we split our data into 10 stratified folds and performed cross-validation on one held-out fold at a time. For all our methods we tune the parameters of our models on a separate validation fold. The overall performance is assessed using Pearson correlation between the set of predicted values and the user-reported score. Results are presented in Table 3.
The results show that both dimensions can be predicted well above chance, with political leaning being easier to predict than engagement. Word2Vec clusters obtain the highest predictive accuracy for political leaning, even though they did not perform as well in the previous classification tasks. For political engagement, political terms and Word2Vec clusters obtain similar predictive accuracy. This result is expected based on the results from Figure 3, which showed how political term usage varies across groups and how it is especially dependent on political engagement. While political terms are very effective at distinguishing between two opposing political groups, they cannot discriminate as well between levels of engagement within the same ideological orientation. Combining all classifiers’ predictions in a linear ensemble obtains best results when compared to each individual category.

### 6.3 Encoding Class Structure

In our previous experiments, we uncovered that certain relationships exist between the seven groups. For example, extreme conservatives and liberals both demonstrate strong political engagement. Therefore, this class structure can be exploited to improve classification performance. To this end, we deploy the sparse graph regularized approach (Argyriou et al., 2007; Zhou et al., 2011) to encode the structure of the seven classes as a graph regularizer in a logistic regression framework.

In particular, we employed a multi-task learning paradigm, where each task is a one-vs-all classification. Multi-task learning (MTL) is a learning paradigm that jointly learns multiple related tasks and can achieve better generalization performance than learning each task individually, especially when presented with insufficient training samples (Liu et al., 2015, 2016b,d). The group structure is encoded into a matrix $R$ which codes the groups which are considered similar. The objective of the sparse graph regularized multi-task learning problem is:

$$
\min_{W,c} \sum_{t=1}^{\tau} \sum_{i=1}^{N} \log(1 + \exp(-Y_{t,i}(W_{t,i}^T X_{t,i} + c_t))) + \gamma \|WR\|_F^2 + \lambda \|W\|_1,
$$

where $\tau$ is the number of tasks, $|N|$ the number of samples, $X$ the feature matrix, $Y$ the outcome matrix, $W_{t,i}$ and $c_t$ is the model for task $t$ and $R$ is the structure matrix.

We define three $R$ matrices: (1) codes that groups with similar political engagement are similar (i.e. 1–7, 2–6, 3–5); (2) codes that groups from each ideological side are similar (i.e. 1–2, 1–3, 2–3, 5–6, 5–7, 6–7); (3) learnt from the data. Results are presented in Table 4. Regular logistic regression performs slightly better than the majority class baseline, which demonstrates that the 7-class classification is a very hard problem although most miss-classifications are within one ideology point. The graph regularization (GR) improves the classification performance over logistic regression (LR) in all cases, with political leaning based matrix (GR–Engagement) obtaining 2% in accuracy higher than the political engagement one (GR–Engagement) and the learnt matrix (GR–Learnt) obtaining best results.

### 7 Conclusions

This study analyzed user-level political ideology through Twitter posts. In contrast to previous work, we made use of a novel data set where fine-grained user political ideology labels are obtained through surveys as opposed to binary self-reports. We showed that users in our data set are far less...
likely to post about politics and real-world fine-grained political ideology prediction is harder and more nuanced than previously reported. We analyzed language differences between the ideological groups and uncovered a dimension of political engagement separate from political leaning.

Our work has implications for pollsters or marketers, who are most interested to identify and persuade moderate users. With respect to political conclusions, researchers commonly conceptualize ideology as a single, left-right dimension similar to what we observe in the U.S. Congress (Ansolabehere et al., 2008; Bafumi and Herron, 2010). Our results suggest a different direction: self-reported political extremity is more an indication of political engagement than of ideological self-placement (Abramowitz, 2010). In fact, only self-reported extremists appear to devote much of their Twitter activity to politics at all.

While our study focused solely on text posted by the user, follow-up work can use other modalities such as images or social network analysis to improve prediction performance. In addition, our work on user-level modeling can be integrated with work on message-level political bias to study how this is revealed across users with various levels of engagement. Another direction of future study will look at political ideology prediction in other countries and cultures, where ideology has different or multiple dimensions.

Acknowledgments

The authors acknowledge the support of the Templeton Religion Trust, grant TRT-0048. We wish to thank Prof. David S. Rosenblum for supporting the research visit of Ye Liu.

References


