"All I know about politics is what I read in Twitter": Weakly Supervised Models for Extracting Politicians' Stances From Twitter

Kristen Johnson
Purdue University
West Lafayette, IN 47907, USA
john1187@purdue.edu

Dan Goldwasser
Purdue University
West Lafayette, IN 47907, USA
dgoldwas@purdue.edu

Abstract

Politicians often use Twitter to express their beliefs, stances on current political issues, and reactions concerning national and international events. Since politicians are scrutinized for what they choose or neglect to say, they craft their statements carefully. Thus despite the limited length of tweets, their content is highly indicative of a politician's stances. We present a weakly supervised method for understanding the stances held by political figures, on a wide array of issues, by analyzing how issues are framed in their tweets and their temporal activity patterns. We combine these components into a global model which collectively infers the most likely stance and agreement patterns, with respective accuracies of 89.84% and 87.76% on average.

1 Introduction

Recently the popularity of traditional media outlets such as television and printed press has decreased, causing politicians to turn their attention to social media outlets, which allow them to directly access the public, express their beliefs, and react to current events. This trend emerged during the 2008 U.S. presidential election campaign and has since moved to the mainstream – in the 2016 campaign, all candidates employ social media platforms. One of the most notable examples of this trend is the micro-blogging outlet Twitter, which unlike its predecessors, requires candidates to compress their ideas, political stances, and reactions into 140 character long tweets. As a result, candidates have to cleverly choose how to frame controversial issues, as well as react to events and each other (Mejova et al., 2013; Tumasjan et al., 2010).

In this work we present a novel approach for modeling the micro-blogging activity of presidential candidates and other prominent politicians. We look into two aspects of the problem, *stance* prediction over a wide array of issues, as well as *agreement and disagreement* patterns between politicians over these issues. While the two aspects are related, we argue they capture different information, as identifying agreement patterns reveals alliances and rivalries between candidates, across and inside their party. We show that understanding the political discourse on micro-blogs requires modeling both the content of posted messages as well as the social context in which they are generated, and suggest a joint model capturing both aspects. Converse to other works predicting stance per individual tweet (SemEval, 2016), we use the *overall* Twitter behavior to predict a *politician's* stance on an issue. We argue that these settings are better suited for the political arena on Twitter. Given the 140 characters limit, the stance relevance of a tweet is not independent of the social context in which it was generated. In an extreme case, even the lack of Twitter activity on certain topics can be indicative of a stance.

For example, consider the issue of gun control. Figure 1 shows three topic-related tweets by three politicians. To correctly identify the stance taken by each of the politicians, our model must combine three aspects. First, the relevance of these tweets to the question can be identified using *issue* indicators (marked in green). Second, the similarity between the stances taken by two of the three politicians can be identified by observing how the issue is *framed* (marked in yellow). Framing issues in order to create bias towards their stance is a tool often used by politicians to contextualize the discussion (Tsur et al., 2015; Card et al., 2015). In this example, tweets (1) and (3) frame the issue of gun control as a matter of safety, while (2) frames it as an issue related to personal freedom, thus revealing the agreement and disagreement patterns between them. Finally, we note the strong negative sentiment of tweet (1). Notice that each aspect individually might not contain sufficient information for correct classification, but combining all three, by propagating the stance bias (derived from analyzing the negative sentiment of (1)) to politicians likely to hold similar or opposing views (derived from frame analysis), leads to a more reliable prediction.

(1) **@HillaryClinton** *We need to keep guns out of the hands of domestic abusers and convicted stalkers*.

(2) **@realDonaldTrump** *Politicians are trying to chip away at the 2nd Amendment. I won't let them take away our guns!*

(3) **@SenSanders** *We need sensible gun-control legislation which prevents guns from being used by people who should not have them* .

Figure 1: Tweets on the issue of gun control, highlighting issue indicators in green and different frame indicators in yellow.

Given the dynamic nature of this domain, we design our approach to use minimal supervision and naturally adapt to new issues. Our model builds on several weakly supervised local learners that use a small seed set of issue and frame indicators to characterize the stance of tweets (based on lexical heuristics (O'Connor et al., 2010) and framing dimensions (Card et al., 2015)) and activity statistics which capture temporally similar patterns between politicians' Twitter activity. Our final model represents agreement and stance bias by combining these weak models into a weakly supervised joint model through Probabilistic Soft Logic (PSL), a recently introduced probabilistic modeling framework (Bach et al., 2013). PSL combines these aspects declaratively by specifying high level rules over a relational representation of the politicians' activities (exemplified in Figure 2), which is further compiled into a graphical model called a hinge-loss Markov random field (Bach et al., 2013), and used to make predictions about stance and agreement between politicians.

We analyze the Twitter activity of 32 prominent U.S. politicians, some of which were candidates for the U.S. 2016 presidential election. We collected their recent tweets and stances on 16 different issues, which were used for evaluation purposes. Our experiments demonstrate the effectiveness of our global modeling approach, which outperforms both a supervised version of the global model and the weak learners that provided the initial supervision.

2 Related Work

To the best of our knowledge this is the first work predicting *politicians' stances using Twitter data, based on content, frames, and temporal activity*. Several works (Sridhar et al., 2015; Hasan and Ng, 2014; Abu-Jbara et al., 2013; Walker et al., 2012; Abbott et al., 2011; Somasundaran and Wiebe, 2010; Somasundaran and Wiebe, 2009) have studied mining opinions and predicting stances in online debate forum data, exploiting argument and threaded conversation structures, both of which are not present in short Twitter data. Social interaction and group structure has also been explored (Sridhar et al., 2015; Abu-Jbara et al., 2013; West et al., 2014). Works focusing on inferring signed social networks (West et al., 2014), stance classification (Sridhar et al., 2015), social group modeling (Huang et al., 2012), and PSL collective classification (Bach et al., 2015) are closest to our work, but these typically operate in supervised settings. Conversely, we use PSL *without direct supervision*, to assign *soft* values (0 to 1 inclusive) to output variables, rather than Markov Logic Networks, which assign *hard* (0 or 1) values to model variables and incur heavier inference time computational cost.

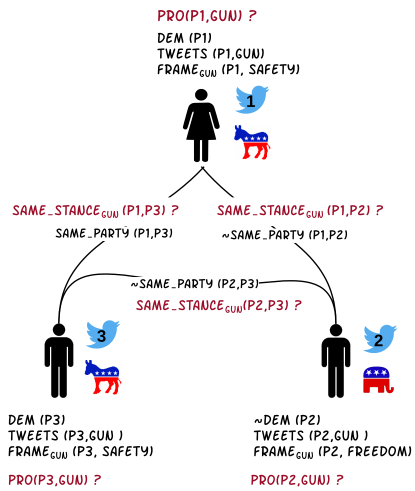


Figure 2: Relational Representation of Politicians' Twitter Activity. P1, P2, and P3 represent 3 different politicians. GUN refers to the issue of gun control; SAFETY and FREEDOM refer to different frames. Prediction target predicates are marked in red.

In recent years there has been a growing interest in analyzing political discourse. Several previous works have explored topic framing (Tsur et al., 2015; Card et al., 2015; Baumer et al., 2015) of public statements, congressional speeches, and news articles. Other works focus on identifying and measuring political ideologies (Iyyer et al., 2014; Bamman and Smith, 2015; Sim et al., 2013) and policies (Gerrish and Blei, 2012; Nguyen et al., 2015). To the best of our knowledge, this work is also the *first attempt to analyze issue framing in Twitter data*. To do so we used the frame guidelines developed by (Boydstun et al., 2014). Issue framing is related to both analyzing biased language (Greene and Resnik, 2009; Recasens et al., 2013) and subjectivity (Wiebe et al., 2004). Unsupervised and weakly supervised models of Twitter data for several various tasks have been suggested, such as user profile extraction (Li et al., 2014b), life event extraction (Li et al., 2014a), and conversation modeling (Ritter et al., 2010). Further, Eisenstein (2013) discusses methods for dealing with the unique language used in micro-blogging platforms.

Analyzing political tweets has also attracted considerable interest. Recently, SemEval Task 6 (SemEval, 2016) aimed to detect stance of *individual tweets*. In contrast to this task, as well as most related work on stance prediction (e.g., those mentioned above), we *do not assume that each tweet expresses a stance*. Instead, we investigate how a politician's overall Twitter behavior, as represented by combined content and temporal indicators, is indicative of a stance (e.g., also capturing when politicians *fail to tweet about a topic*). Predicting political affiliation and other characteristics of Twitter users has been explored (Volkova et al., 2015; Volkova et al., 2014; Conover et al., 2011). Other works have focused on sentiment analysis (Pla and Hurtado, 2014; Bakliwal et al., 2013), predicting ideology (Djemili et al., 2014), automatic polls based on Twitter sentiment and political forecasting using Twitter (Bermingham and Smeaton, 2011; O'Connor et al., 2010; Tumasjan et al., 2010), as well as uses of distant supervision (Marchetti-Bowick and Chambers, 2012).

3 Data and Problem Setting

REPUBLICAN POLITICIANS	DEMOCRATIC POLITICIANS
Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Carly Fiorina, Lindsey Graham, Mike Huckabee, Bobby Jindal, John Kasich, George Pataki, Rand Paul, Rick Perry, Marco Rubio, Rick Santorum, Donald Trump, Scott Walker	Joe Biden, Lincoln Chafee, Hillary Clinton, Kirsten Gillibrand, John Kerry, Ben Lujan, Ed Markey, Martin O'Malley, Nancy Pelosi, Harry Reid, Bernie Sanders, Chuck Schumer, Jon Tester, Mark Warner, Elizabeth Warren, Jim Webb

Table 1: Politicians tracked in this study.

ISSUE	QUESTION
ABORTION	*Do you support abortion?*
ACA	*Do you support the Patient Protection and Affordable Care Act (Obamacare)?*
CONFEDERATE	*Should the federal government allow states to fly the confederate flag?*
DRUGS	*Do you support the legalization of Marijuana?*
ENVIRONMENT	*Should the federal government continue to give tax credits and subsidies to the wind power industry?*
GUNS	*Do you support increased gun control?*
IMMIGRATION	*Do you support stronger measures to increase our border security?*
IRAN	*Should the U.S. conduct targeted airstrikes on Irans nuclear weapons facilities?*
ISIS	*Should the U.S. formally declare war on ISIS?*
MARRIAGE	*Do you support the legalization of same sex marriage?*
NSA	*Do you support the Patriot Act?*
PAY	*Should employers be required to pay men and women, who perform the same work, the same salary?*
RELIGION	*Should a business, based on religious beliefs, be able to deny service to a customer?*
SOCIAL SECURITY	*Should the government raise the retirement age for Social Security?*
STUDENT	*Would you support increasing taxes on the rich in order to reduce interest rates for student loans?*
TPP	*Do you support the Trans-Pacific Partnership?*

Table 2: Issues taken from `ISideWith.com` and their corresponding Yes/No questions.

Data Collection and Pre-Processing: We collected tweets post 2009 for 32 politicians, the 16 Republicans and 16 Democrats listed in Table 1. Our initial goal was to compare politicians participating in the 2016 U.S. presidential election (16 Republicans and 5 Democrats). To increase representation of Democrats, we collected tweets of Democrats who hold leadership roles within their party. We focused on well known politicians because they tend to tweet with a focus on national rather than local (district/state) events. For all 32 politicians we have a total of 99,161 tweets, with an average of 3,000 per person. There are 39,353 Democrat and 59,808 Republican tweets. All data, including tweets, keywords, and PSL scripts will be made available at: `www.***.***`.

Using tweets from both parties, we compiled a set of frequently appearing keywords for each issue, with an average of seven keywords per issue. A Python script then used these preselected keywords to filter all tweets, keeping only those that represent our 16 political issues of interest (shown in Table 2), and automatically eliminating all irrelevant tweets (e.g., those about personal issues, campaigning, duplicates, and non-English tweets). We intentionally used a small set of keywords to avoid overselection, i.e., avoiding tweets about praying for a friend's *health* but retaining tweets discussing *health care*.

Annotating Stances and Agreement: Using `ISideWith.com`, a popular website that provides users with how strongly their opinions match to politicians based on their answers to a series of 58 questions, we chose 16 of these as issues for our prediction goals. For each of these 16 issues, at least 15 (with an average of 26) of the 32 politicians have tweeted on that issue; for the remaining issues, we found fewer than half the politicians or none tweeted about that issue. These issues range over common policies including domestic and foreign policy, economy, education, environment, health care, immigration, and social issues. `ISideWith.com` uses a range of yes/no answers in their questions and provides proof of the politician's stance on that issue, if available, through public information such as quotes or voting records. When unavailable, the site assigns a politician an answer based on party lines or has no answer.

Since we use the stances as the ground truth for evaluating our prediction, all politicians with unavailable answers or those not listed on the site were manually annotated via online searches of popular newspapers, political channels, and voting records. It is important to note that `ISideWith.com` does not contain answers to all questions for all politicians, especially those that are less popular. Our weakly supervised approach requires no prior knowledge of the politician's stance and therefore generalizes to situations where such information is unavailable. Instead, our model only uses keywords describing topics and frames, Twitter behavior patterns, and party information, all of which is easily attainable and adaptable for new domains (e.g., different keywords to capture important issues in another country).

Predicting Stance and Agreement: The previously collected stances represent the ground truth of whether a politician is for or against an issue. Based on these we define two target predicates using PSL notation (see Section 5.1) to capture the desired output as soft truth assignments to these predicates. The first predicate, PRO(P1, ISSUE) captures a positive stance by politician P1, on an ISSUE. Consequently, a negative stance would be captured by its negation: ¬PRO(P1, ISSUE). The second target predicate, SAMESTANCE$_I$(P1, P2) classifies if two politicians share a stance for a given issue, i.e., if both are for or against an issue, where *I* represents 1 of the 16 issues being investigated. Although the two predicates are clearly inter-dependent, we chose to model them as separate predicates since they can depend on different Twitter behavioral and content cues. Indeed, given the short and context-free style of Twitter we can often find indicators of politicians holding similar stances, *without* clear specification for which stance they actually hold.

4 Local Models of Twitter Activity

Our model builds on a collection of weakly supervised local models which provide an initial bias when learning the parameters of the global PSL model. The local models capture similarity between tweet content and temporal activity patterns of users' timelines as well as stance bias.

4.1 Issue

To capture which issues politicians are tweeting about, we build a keyword based heuristic, similar to the approach described in (O'Connor et al., 2010). Each issue is associated with a small set of keywords, as described in the previous section. The keywords appearing in a given tweet may be mutually exclusive, such as those concerning Iran or Environment; however, some may fall under multiple issues at once (e.g., *religion* may indicate the tweet refers to ISIS, Religion, or Marriage). Tweets are classified as related to a certain issue based on the majority of matching keywords, with rare cases of ties manually resolved. The output of this classifier is all of the issue-related tweets of a politician, which are used as input for the PSL predicate TWEETS(P1, ISSUE), a binary predicate which indicates if that politician has tweeted about the issue or not.

4.2 Sentiment Analysis

The sentiment of a tweet can help expose a politician's stance on a certain issue. We use OpinionFinder 2.0 (Wilson et al., 2005) to label each politician's issue related tweets as positive, negative, or neutral. We observed, however, that for all politicians, a majority of tweets will be labeled as neutral. This may be caused by the difficulty of labeling sentiment for Twitter data. If this results with a politician having no positive or negative tweets, they are assigned their party majority sentiment assignment for that issue. This output is used as input to the PSL predicates TWEETPOS(P1, ISSUE) and TWEETNEG(P1, ISSUE).

4.3 Agreement and Disagreement

To determine how well tweet content similarity can capture stance agreement, we computed the pairwise cosine similarity between all of the politicians. Unfortunately, the use of similar words per issue resulted in most politicians being grouped together, even across different parties. To overcome this, we compute the *frequency* of similar words within tweets about each issue. For each issue, all of a politician's tweets are aggregated and the frequency of each word is compared to all other politicians' word frequencies. Politicians, P1 and P2, are considered to have a similar LOCALSAMESTANCE$_I$(P1, P2) if their frequency counts per word for an issue are within the same range.

4.4 Temporal Activity Patterns

We observed from reading Twitter feeds that most politicians will comment on an event the day it happens. For general issues, politicians comment as frequently as desired to express their support or lack thereof for that particular issue. For example, Rand Paul tweeted daily in opposition of the NSA during his filibuster of the Patriot Act renewal. Conversely Hillary Clinton has no tweets concerning the NSA or Patriot Act. To capture patterns between politicians, we align their timelines based on days where they have tweeted about an issue. When two or more politicians tweet about the same issue on the same day, they are considered to have similar temporal activity, which may indicate stance agreement. This information is used as input for our PSL predicate SAMETEMPORALACTIVITY$_I$(P1, P2).

4.5 Political Frames

We follow the intuition that the way politicians choose to contextualize their tweets on an issue is strongly indicative of their stance on that issue. To investigate this, we compiled a list of unique keywords for each political framing dimension as described in Boydstun et al. (2014) and Card et al. (2015). We again use the keyword matching approach described in Section 4.1 to classify all tweets into a political frame. As noted in Card et al. (2015), some tweets may fall into multiple frames. After all tweets are labeled, we sum over the total number of each frame type and use the frame with the maximum count and second largest count as that politician's frames for that issue. We use the top two frames because for nearly all politicians a majority of their issue related tweets will fall into two frames. In the event of a tie we assign the frame that appears most frequently within that politician's party. These frames are used as input to the PSL predicate FRAME(P1, ISSUE).

5 Global Models of Twitter Activity

Finally, we use PSL to tie together all local models into a joint global model. As shown by our baseline measurements in Section 6, local information alone is not strong enough to quantify stance or agreement for politicians. Using PSL, we are able to build connections between each local model and thus increase the overall accuracy of each global model's prediction. In addition to the PSL predicates representing the target output (PRO and SAMESTANCE$_I$) [1] and local models (as defined in Section 4), we also use directly observed information: party affiliation, denoted DEM(P1) for Democrat and $\neg$DEM(P1) for Republican, and SAMEPARTY(P1, P2) to denote if two politicians belong to the same party.

5.1 Global Modeling using PSL

PSL is a recent declarative language for specifying weighted first-order logic rules. A PSL model is specified using a set of weighted logical formulas, which are compiled into a special class of graphical model, called a hinge-loss MRF, defining a probability distribution over the possible continuous value assignments to the model's random variables and allowing the model to scale easily (Bach et al., 2015). The defined probability density function has the form:

$$P(\mathbf{Y} \mid \mathbf{X}) = \frac{1}{Z} \exp\left(-\sum_{r=1}^{M} \lambda_r \phi_r(\mathbf{Y}, \mathbf{X})\right)$$

where λ is the weight vector, Z is a normalization constant, and

$$\phi_r(\mathbf{Y}, \mathbf{X}) = (\max\{l_r(\mathbf{Y}, \mathbf{X}), 0\})^{\rho_r}$$

is the hinge-loss potential corresponding to the instantiation of a rule, specified by a linear function l_r, and an optional exponent $\rho_r \in 1, 2$. The weights of the rules are learned using maximum-likelihood estimation, which in our weakly supervised setting was estimated using the Expectation-Maximization algorithm. For more details we refer the reader to Bach et al. (2015).

Specified PSL rules have the form:

[1] In a supervised setting, jointly modeling the 2 target predicates can improve performance. Experiments using this approach yielded improvement in performance *and* a more complex model containing more parameters, resulting in slower inference.

$$\lambda_1 : P_1(x) \wedge P_2(x,y) \rightarrow P_3(y), \qquad \lambda_2 : P_1(x) \wedge P_4(x,y) \rightarrow \neg P_3(y)$$

where P_1, P_2, P_3, P_4 are predicates, and x, y are variables. Each rule is associated with a weight λ, which indicates its importance in the model. Given concrete constants *a, b* respectively instantiating the variables x, y, the mapping of the model's atoms to soft [0,1] assignments will be determined by the weights assigned to each one of the rules. For example, if $\lambda_1 > \lambda_2$, the model will prefer $P_3(b)$ to its negation. This contrasts with "classical" or other probabilistic logical models in which rules are strictly true or false. In our domain, the constant symbols correspond to politicians and predicates to: party affiliation, Twitter activity, and similarity between political figures based on Twitter behaviors.

5.2 Baseline: Using Local Classifiers Directly

Previous works exploring stance classification typically predict stance based on a single piece of text (e.g., debates, forum posts, tweets) in a supervised setting, making it difficult to directly compare to our approach. To facilitate comparison, we implement a baseline model and a supervised version of our best performing model (described in 5.5), which, as expected, have a weaker performance than our models. The baseline model does not take advantage of the global modeling framework, but instead learns weights over the rules listed in Table 3 which directly map the output of local noisy classifiers to PSL target predicates.

5.3 Model 1: Agreement with Party Lines

The observation that politicians tend to vote with their political party on most issues is the basis of our initial assumptions in Model 1. This tendency is encoded via the PSL rules listed in Table 4 which aim to capture party based agreement. For some issues we initially assume Democrats (DEM) are for an issue, while Republicans (¬DEM) are against that issue, or vice versa. In the latter case, the rules of the model would change accordingly, e.g. the second rule would become ¬DEM(P1) →PRO(P1, ISSUE), and likewise for all other rules. Similarly, if two politicians are in the same party, we expect them to have the same stance, or agree, on an issue. Though this is a strong initial assumption, the model can incorporate other indicators to overcome this bias when necessary. For all PSL rules, the reverse also holds, e.g., if two politicians are not in the same party, we expect them to have different stances.

PSL Rules: LOCAL BASELINE MODEL (LB)
LOCALSAMESTANCE$_I$(P1, P2) →SAMESTANCE$_I$(P1, P2)
¬LOCALSAMESTANCE$_I$(P1, P2) → ¬SAMESTANCE$_I$(P1, P2)
TWEETS(P1,ISSUE) ∧TWEETPOS(P1,ISSUE) → PRO(P1, ISSUE)
TWEETS(P1,ISSUE) ∧TWEETNEG(P1,ISSUE) → ¬PRO(P1, ISSUE)

Table 3: Subset of PSL Rules Used in Local Baseline.

PSL Rules: MODEL 1 (M1)
SAMEPARTY(P1, P2) →SAMESTANCE$_I$(P1, P2)
DEM(P1) →PRO(P1, ISSUE)
¬DEM(P1) →¬PRO(P1, ISSUE)
SAMEPARTY(P1, P2) ∧Dem(P1) →PRO(P2, ISSUE)
SAMEPARTY(P1, P2) ∧¬Dem(P1) →¬PRO(P2, ISSUE)
SAMEPARTY(P1,P2) ∧Pro(P1, Issue) ∧DEM(P1) → PRO(P2, ISSUE)

Table 4: Subset of PSL Rules Used in Model 1.

5.4 Model 2: Politicians' Twitter Activity

Model 2 builds upon the initial party line bias of Model 1. In addition to political party based information, we also include representations of the politician's Twitter activity, as shown in Table 5. This includes whether or not a politician tweets about an issue (TWEETS) as well as the sentiment of the tweets as determined in Section 4.2. The predicate TWEETPOS models if a politician tweets positively on the issue, whereas TWEETNEG models negative sentiment. Two sentiment predicates are used instead of the negation of TWEETPOS, which would cause all politicians for which there are no tweets, and hence no sentiment, on that issue to also be considered.

PSL Rules: MODEL 2 (M2)
TWEETS(P1, ISSUE) ∧DEM(P1) →PRO(P1, ISSUE)
TWEETS(P1, ISSUE) ∧¬DEM(P1) →¬PRO(P1, ISSUE)
TWEETS(P1, ISSUE) ∧TWEETS(P2, ISSUE) ∧SAMEPARTY(P1, P2) →SAMESTANCE$_I$(P1, P2)
TWEETPOS(P1, ISSUE) ∧TWEETPOS(P2, ISSUE) →SAMESTANCE$_I$(P1, P2)
TWEETPOS(P1, ISSUE) ∧TWEETNEG(P2, ISSUE) →¬SAMESTANCE$_I$(P1, P2)

Table 5: Subset of PSL Rules Used in Model 2. Corresponding TWEETNEG rules are omitted due to space.

PSL Rules: Model 3 (M3)
LocalSameStance$_I$(P1, P2) $\wedge$Pro(P1, Issue) $\rightarrow$Pro(P2, Issue)
SameTemporalActivity$_I$(P1, P2) $\wedge$SameParty(P1, P2) $\rightarrow$ SameStance$_I$(P1, P2)
SameTemporalActivity$_I$(P1, P2) $\wedge$Frame(P1,Issue) $\wedge$Frame(P2, Issue) $\rightarrow$SameStance$_I$(P1, P2)
Frame(P1, Issue) $\wedge$Frame(P2, Issue) $\rightarrow$SameStance$_I$(P1, P2)
Frame(P1, Issue) $\wedge$Frame(P2, Issue) $\wedge$SameParty(P1,P2) $\rightarrow$SameStance$_I$(P1, P2)

Table 6: Subset of PSL Rules Used in Model 3.

Issue	Stance				Agreement			
	LB	M 1	M 2	M 3	LB	M 1	M 2	M 3
Abortion	81.25	96.88	96.88	96.88	49.31	93.75	93.75	95.36
ACA	96.88	100	100	100	51.61	100	100	100
Confederate	34.38	78.12	87.5	84.38	51.31	69.6	77.7	80.18
Drugs	87.5	78.12	96.88	88.88	50.42	63.6	84.07	84.07
Environment	53.12	78.12	78.13	81.08	45.16	68.75	65.59	69.28
Guns	93.75	93.75	93.75	93.75	48.59	68.54	99.59	99.59
Immigration	37.5	81.25	81.25	86.36	53.62	68.55	69.06	69.56
Iran	84.38	65.62	65.63	84.38	35.57	79.73	100	100
ISIS	40.32	76.28	93.75	93.75	59.68	76.28	76.28	90.04
Marriage	62.5	90.62	90.62	90.62	50.57	87.12	87.43	87.43
NSA	37.5	53.12	53.12	61.54	34.15	49.2	56.66	59.65
Pay	84.38	84.38	90.62	89.47	64.30	72.92	80.31	74.31
Religion	75	68.75	81.25	81.25	47.62	86.24	76.46	79.44
Social Security	28.12	78.12	78.13	78.13	53.76	73.25	90.03	90.88
Student	93.75	96.88	96.88	96.88	51.61	100	100	100
TPP	62.5	62.5	62.5	62.5	45.43	48.39	54.64	65.32

Table 7: Stance and Agreement Accuracy by Issue. LB uses weak local models, M1 represents party line agreement, M2 adds Twitter activity, M3 adds higher level Twitter behaviors, and SM represents the supervised version of M3 for comparison.

5.5 Model 3: Politicians' Agreement Patterns

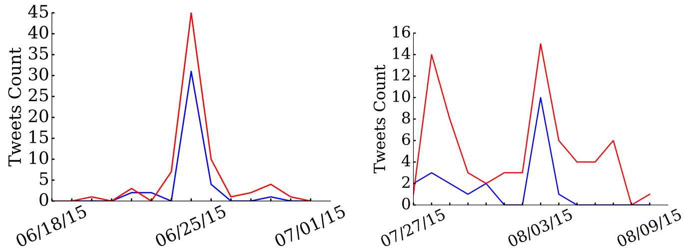
Table 6 presents a subset of the rules used in Model 3 to incorporate higher level Twitter information into the model. Our intuition is that politicians who have similar tweets would also have similar stances on issues, which we represent with the predicate LocalSameStance$_I$. SameTemporalActivity represents the idea that if politicians tweet on an issue around the same time range then they also share a stance for that issue. Finally, Frame indicates the frame used by that politician for different issues. More details on these predicates are in Sections 4.3, 4.4, and 4.5, respectively. The incorporation of these rules allows Model 3 to overcome Model 2 inconsistencies between stance and sentiment (e.g., when someone is attacking their opposition).

6 Experiments

Experimental Settings: As described in Section 4, the data generated from the local models is used as input to the PSL models. Stances collected in Section 3 are used as the ground truth for evaluation of the results of the PSL models. We initialize Model 1, as described in Section 5.3, using knowledge of the politician's party affiliation. Model 2 builds upon Model 1 by incorporating the results of the issue and sentiment analysis local models, as described in Sections 4.1 and 4.2 respectively. Model 3 combines all previous models with higher level knowledge of Twitter activity: tweet agreement (Section 4.3), temporal activity (Section 4.4), and frames (Section 4.5). We implement our PSL models to have an initial bias that candidates do not share a stance and are against an issue. Finally, to provide a comparison to a supervised approach, we implement a supervised version of Model 3 which is trained on 80% of politicians and tested on the remaining 20%, using 5-fold cross-validation.

Experimental Results By Issue: Table 7 presents the results of using our three proposed PSL models. Local Baseline (LB) refers to using only the weak local models for prediction with no additional information about party affiliation. We observe that for prediction of stance (Pro) LB performs better than random chance in 12 of 16 issues; for prediction of agreement (SameStance$_I$), LB performs much lower overall, with only 10 of 16 issues predicted above chance.

Using Model 1 (M1), we improve stance prediction accuracy for 10 of the issues and agreement accuracy for all issues except TPP. This is likely due to the fact that M1 incorporates party information and the issue of TPP is the most heavily divided within and across parties, with 8 Republicans and 4 Democrats in support of TPP and 8 Republicans and 12 Democrats opposed. Model 2 (M2) further



a ACA b Abortion

Figure 3: Temporal Twitter Activity by Party. Republican (red) and Democrat (blue) event based temporal overlaps.

improves the stance and agreement predictions for an additional 6 and 11 issues, respectively. The most interesting cases to note here are those of ISIS and Iran - using just M2, we are able to achieve 100% accuracy for agreement prediction. Model 3 (M3), the combination of the previous models with higher level Twitter features, increases the stance prediction accuracy of M2 for 7 issues and the agreement accuracy for 5 issues. The final predictions of M3 are significantly improved over the initial LB for all issues, except TPP. Even in cases where M1 and/or M2 lowered the initial baseline result (e.g. stance for Drug or agreement for TPP), the final prediction by M3 is still significantly higher than that of the baseline. Finally, our weakly supervised M3 outperforms the baseline supervised version (SM) on all predictions except ACA and ties for stance prediction on TPP. This is because M3 can consider all relationships across the entire network, while SM cannot. The supervised setting achieves better ACA results because the parties are perfectly split on this case. Both approaches tie on TPP stance prediction because the parties are heavily mixed for this issue, as described previously.

Framing and Temporal Information: As shown in Table 7, performance for *some* issues does not improve in Model 3. Upon investigation, we found that for all issues, except Abortion which improves in agreement, one or both of the top frames for the party are shared across party lines. For example, for ACA both Republicans and Democrats have the *Economic* and *Health and Safety* frames as their top two frames. For Immigration, Democrats have the *Economic* frame, Republicans have the *Morality* frame, and both parties share the *Security and Defense* frame. In addition to similar framing overlap, the Twitter timeline for ACA also exhibits overlap, as shown in Figure 8a. This figure highlights one week before and after the Supreme Court ruling to uphold the ACA. The peak of Twitter activity is the day of the ruling, 6/25/2015. Conversely, Abortion which shares no frames between parties (Democrats frame Abortion with *Constitutionality* and *Health and Safety* frames; Republicans use *Economic* and *Capacity and Resources* frames), exhibits a timeline with greater fluctuation. The peak of Figure 8b is 8/3/2015, which is the day that the budget was passed to include funding for Planned Parenthood. Despite sharing a peak, both parties have different patterns over this time frame, allowing Model 3 to extract enough information to increase agreement prediction accuracy by 2.02%.

7 Conclusion

In this paper we take a first step towards understanding the dynamic micro-blogging behavior of politicians. Though we concentrate on a small set of politicians and issues in this work, this framework can be modified to handle additional politicians or issues, as well as those in other countries, by incorporating appropriate domain knowledge (e.g., replacing party with voting history, using new keywords for different issues in other countries, or changing events such as Supreme Court rulings to Parliament votes), which we leave as future work. Unlike previous works, which tend to focus on one aspect of this complex micro-blogging behavior, we build a holistic model connecting temporal behaviors, party-line bias, and issue frames into a single predictive model used to identify fine-grained policy stances and agreement. Despite having no explicit supervision, and using only intuitive "rules-of-thumb" to bootstrap our global model, our approach results in a strong prediction model which helps shed light on political discourse framing inside and across party lines.

References

Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani, and Joseph King. 2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proc. of the Workshop on Language in Social Media*.

Amjad Abu-Jbara, Ben King, Mona Diab, and Dragomir Radev. 2013. Identifying opinion subgroups in arabic online discussions. In *Proc. of ACL*.

Stephen H. Bach, Bert Huang, Ben London, and Lise Getoor. 2013. Hinge-loss Markov random fields: Convex inference for structured prediction. In *Proc. of UAI*.

Stephen H Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2015. Hinge-loss markov random fields and probabilistic soft logic. *arXiv preprint arXiv:1505.04406*.

Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi, and Mark Hughes. 2013. Sentiment analysis of political tweets: Towards an accurate classifier. In *Proc. of ACL*.

David Bamman and Noah A Smith. 2015. Open extraction of fine-grained political statements. In *Proc. of EMNLP*.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *In Proc. of ACL*.

Adam Bermingham and Alan F Smeaton. 2011. On using twitter to monitor political sentiment and predict election results.

Amber Boydstun, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2014. Tracking the development of media frames within and across policy issues.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proc. of ACL*.

Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Proc. of Privacy, Security, Risk and Trust (PASSAT) and SocialCom*.

Sarah Djemili, Julien Longhi, Claudia Marinica, Dimitris Kotzinos, and Georges-Elia Sarfati. 2014. What does twitter have to say about ideology? In *NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication*.

Jacob Eisenstein. 2013. What to do about bad language on the internet. In *Proc. of NAACL*.

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

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

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proc. of EMNLP*.

Bert Huang, Stephen H. Bach, Eric Norris, Jay Pujara, and Lise Getoor. 2012. Social group modeling with probabilistic soft logic. In *NIPS Workshops*.

Mohit Iyyer, Peter Enns, Jordan L Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proc. of ACL*.

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard H Hovy. 2014a. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proc. of EMNLP*.

Jiwei Li, Alan Ritter, and Eduard H Hovy. 2014b. Weakly supervised user profile extraction from twitter. In *Proc. of ACL*.

Micol Marchetti-Bowick and Nathanael Chambers. 2012. Learning for microblogs with distant supervision: Political forecasting with twitter. In *Proc. of EACL*.

Yelena Mejova, Padmini Srinivasan, and Bob Boynton. 2013. Gop primary season on twitter: popular political sentiment in social media. In *WSDM*.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proc. of ACL*.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proc. of ICWSM*.

Ferran Pla and Lluís F Hurtado. 2014. Political tendency identification in twitter using sentiment analysis techniques. In *Proc. of COLING*.

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

Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of twitter conversations. In *Proc. of NAACL*.

SemEval. 2016. Task 6. `http://alt.qcri.org/semeval2016/task6/`.

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

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proc. of ACL*.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proc. of NAACL Workshops*.

Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proc. of ACL*.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proc. of ACL*.

Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment.

Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring user political preferences from streaming communications. In *Proc. of ACL*.

Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring latent user properties from texts published in social media. In *Proc. of AAAI*.

Marilyn A. Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proc. of NAACL*.

Robert West, Hristo S Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *TACL*.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational linguistics*.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Opinionfinder: A system for subjectivity analysis. In *In Proc. of EMNLP*.