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

Kristen Johnson and **Dan Goldwasser**
Department of Computer Science
Purdue University, West Lafayette, IN 47907
{john1187, dgoldwas}@purdue.edu

Abstract

During the 2016 United States presidential election, politicians have increasingly used Twitter to express their beliefs, stances on current political issues, and reactions concerning national and international events. Given the limited length of tweets and the scrutiny politicians face for what they choose or neglect to say, they must craft and time their tweets carefully. The content and delivery of these tweets is therefore highly indicative of a politician's stances. We present a weakly supervised method for extracting how issues are framed and temporal activity patterns on Twitter for popular politicians and issues of the 2016 election. These behavioral components are combined into a global model which collectively infers the most likely stance and agreement patterns among politicians, with respective accuracies of 86.44% and 84.6% on average.

1 Introduction

The trending decline in popularity of traditional media outlets and continued rise of social media usage emerged in the 2008 U.S. presidential election campaign and has continued to the present 2016 campaign. Social media platforms, such as the microblogging outlet Twitter, allow politicians to directly access the public, express their beliefs, and react to current events. Unlike its traditional media predecessors, Twitter requires politicians to compress their ideas, political stances, and reactions to 140 character long tweets. Consequently, politicians must cleverly choose how to frame controversial issues, as well as how and when to react to each other (Mejova et al., 2013; Tumasjan et al., 2010). Due to this limit, we argue that the stance of a tweet is not independent of the social context in which it was generated. Thus, for accurate predictions these social behaviors must also be modeled.

Converse to previous works which predict stance per individual tweet (SemEval, 2016), we instead present a novel approach better suited to model the dynamic political arena of Twitter, which uses the *overall* Twitter behavior per politician to predict a *politician's* stance on an issue. We explore two aspects of the problem, *stance* prediction over a wide array of issues, as well as stance *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 within their party. In an extreme case, even the lack of Twitter activity on certain issues can be indicative of a stance.

For example, consider the three tweets on the issue of gun control shown in Figure 1. To identify the stance taken by each politician, our model combines both content and behavioral features, accumulated from all of a politician's tweets on that issue. First, the tweet's relevance to an issue can be identified using *issue* indicators (highlighted in green). Second, the similarity between the stances taken by two of the politicians (agreement) can be identified by observing differences in how the issue is *framed* (shown in yellow), a tool often used by politicians to create bias toward a stance and contextualize the discussion (Tsur et al., 2015; Card et al., 2015). Tweets (1) and (3) frame the issue as a matter of safety, while tweet (2) frames it as related to personal freedom, thus revealing the agreement and disagreement patterns between the politicians. Third, we can consider the timing of these tweets, i.e. whether these tweets are posted continually or just around events concerning gun violence. Finally, we can also use sentiment indicators (e.g., the negative sentiment of tweet (1)). Notice that each feature individually might not contain sufficient information for correct classification, but combining all aspects,

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

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

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

Figure 1: Tweets Discussing the Issue of Gun Control. Issue indicators (e.g. guns and gun-control) are highlighted in green and different frame indicators (e.g., domestic abusers or 2nd Amendment) are highlighted in yellow.

by propagating stance bias (e.g. from sentiment) to politicians who hold similar or opposing views (as determined from frame analysis), leads to a more reliable prediction.

Given the dynamic nature of political discourse on Twitter, we design our approach to use minimal supervision and naturally adapt to new issues. We build several weakly supervised local learners, whose only supervision is a small seed set of issue and frame indicators which characterize the stance of tweets (based on lexical heuristics (O'Connor et al., 2010) and framing dimensions (Card et al., 2015)), and Twitter activity statistics which capture temporally similar patterns between politicians. 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 recent probabilistic modeling framework (Bach et al., 2013). The information gained from the weakly supervised local learners is the only supervision used by PSL; the rest of the prediction is completely unsupervised. 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.

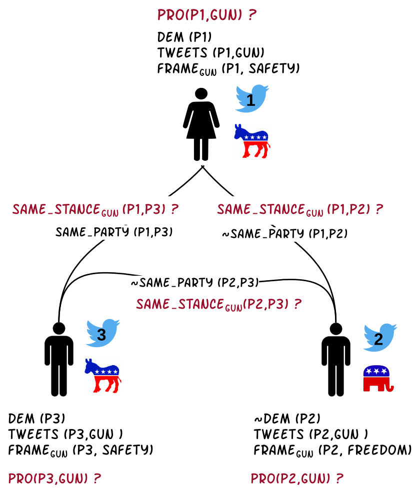


Figure 2: Relational Representation Example of Twitter Activity. P1, P2, and P3 represent 3 different politicians. Prediction target predicates (PRO and SAMESTANCE) are shown in red. Indicators of Twitter content and behavior include: DEM, TWEETS, FRAME$_{GUN}$, SAMEPARTY. GUN refers to the issue of gun control; SAFETY and FREEDOM refer to different frames for the issue.

We analyze the Twitter activity of 32 prominent U.S. politicians, including those who were U.S. 2016 presidential candidates, on 16 different issues. Our experiments demonstrate the effectiveness of our global modeling approach, which outperforms both the weak learners that provide the initial supervision and a supervised text based baseline. Our results show that understanding political discourse on Twitter requires modeling not only the word content of tweets but the social behavior associated with those tweets as well.

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 by exploiting argument and threaded conversation structures, both of which are not always present in short Twitter data[1]. 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.

In recent years there has been a growing interest in analyzing political discourse in both traditional

[1] In our data set, there are few "@" mentions or retweet examples forming a conversation, thus we do not have access to argument or conversation structures for analysis.

and social media. 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; Lewenberg et al., 2016) and policies (Gerrish and Blei, 2012; Nguyen et al., 2015; Grimmer, 2010). To the best of our knowledge, this work is also the *first attempt to analyze issue framing in Twitter data*. To do so we use 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).

Concerning Twitter specifically, analysis of users and political tweets has attracted considerable attention. 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 microblogging platforms.

Recently, SemEval Task 6 (SemEval, 2016) aimed to detect the 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), analyzing types of tweets and Twitter network effects around political events (Maireder and Ausserhofer, 2013), 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 applications of distant supervision (Marchetti-Bowick and Chambers, 2012).

3 Data and Problem Setting

3.1 Data Collection

We collected tweets for 32 politicians, the 16 Republicans (all 2016 presidential candidates) and 16 Democrats (5 of which were candidates) listed in Table 1. Our initial goal was to compare politicians participating in the 2016 U.S. presidential election. To increase representation of Democrats, we collected tweets of Democrats who hold leadership roles within their party, because more well known politicians tend to focus their tweets on national rather than local (district/state) events. For all 32 politicians we have a total of 99,161 tweets: 39,353 Democrat and 59,808 Republican[2].

Based on tweet availability and politician coverage[3], we chose 16 issues (shown in Table 2) derived from the 58 questions used by `ISideWith.com` to match a user to politicians based on their responses as our stance prediction goals. These issues range over common policies including domestic and foreign policy, economy, education, environment, health care, immigration, and social issues.

Republicans	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
Democrats	Candidates: Lincoln Chafee, Hillary Clinton, Martin O'Malley, Bernie Sanders, Jim Webb
	Non-candidates: Joe Biden, Kirsten Gillibrand, John Kerry, Ben Lujan, Ed Markey, Nancy Pelosi, Harry Reid, Chuck Schumer, Jon Tester, Mark Warner, Elizabeth Warren

Table 1: Politicians Tracked in This Study. All Republicans were 2016 presidential candidates. Democrats are divided by whether or not they ran as a candidate.

[2]Our Twitter data set, keywords, and PSL scripts are available at: `purduenlp.cs.purdue.edu/projects/politicaltwitter`.

[3]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 (or none) of the politicians tweeted about that issue.

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: Sixteen Political Issues Used in This Study. Issues and their corresponding Yes/No questions were taken from `ISideWith.com`.

3.2 Data Pre-Processing

Using all tweets, we compiled a set of frequent keywords (an average of 7) for each issue. This set is small to avoid overselection, i.e., avoiding tweets about praying for a friend's *health* but keeping tweets discussing *health care*. Via Python scripts, these keywords are used to retain tweets related to the 16 issues shown in Table 2, while eliminating all irrelevant tweets (e.g., those about personal issues, campaigning, duplicates, and non-English tweets).

`ISideWith.com` uses a range of yes/no answers to their questions and provides proof (through quotes or voting records) of a politician's stance on that issue, *if available*. When unavailable, the site assigns an answer based on party lines or often provides no answer. Also, less popular politicians are not featured on the site. For these cases, we manually annotated the stance using online searches of newspapers or voting records. These stances are only used for evaluation of our predictions. Our weakly supervised approach requires *no* prior knowledge of the politician's stance, allowing it to generalize to situations such as these, where stance information is unavailable.

3.3 Prediction Goals

The 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. A negative stance would be captured by its negation: $\neg$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 of interest. 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. 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

The only supervision required by our method consists of the keywords describing issues and frames, Twitter behavior patterns, and party affiliation, all of which is easily attainable and adaptable for new domains (e.g., different keywords to capture issues of another country). The weakly supervised local models described in this section capture similarities between tweet content and temporal activity patterns of users' timelines, as well as stance bias, and are used to provide the initial bias when learning the parameters of the otherwise unsupervised global PSL model.

4.1 Issue of Tweets

To capture which issues politicians are tweeting about, we used a keyword based heuristic, similar to the approach described in O'Connor et al. (2010), where each issue is associated with a small set of pre-selected keywords (as described previously). The keywords appearing in a given tweet may be mutually exclusive (e.g., *fracking* for Environment will not appear in tweets discussing other issues); 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 relating 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 of Tweets

The sentiment of a tweet can indicate a politician's stance on a certain issue. OpinionFinder 2.0 (Wilson et al., 2005) is used 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. When this results with a politician having no positive or negative tweets, they are assigned their party's majority sentiment for that issue. The majority sentiment of a party is calculated by running all politicians' tweets through OpinionFinder and using whichever sentiment (positive or negative) is assigned the most per party. This output is used as input to the PSL predicates TWEETPOS(P1, ISSUE) and TWEETNEG(P1, ISSUE).

4.3 Content Agreement and Disagreement Patterns

We expect politicians that have a similar stance on an issue to use similar words in their tweets. To determine how well tweet content similarity captures agreement between politicians, we computed the pair-wise cosine similarity between all of the politicians' words used in tweets per issue. However, the use of similar words per issue resulted in most politicians being grouped together, even across different parties. To overcome this, we calculated the *frequency* of similar words within tweets (per issue). For each issue, all of a politician's words from 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 shared word of an issue are within the same range. For this study, we used a window of 10 (i.e., if the frequency count of a word is 30, then a count of 20 to 40 would be considered similar) to ensure that politicians who briefly mention an issue are not considered equivalent to those who discuss it more frequently.

4.4 Temporal Activity Patterns

We observed from reading Twitter feeds that most politicians tweet about an event the day it happens. However, for general issues, politicians will 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 agreement 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 to the predicate $\text{SAMETEMPORALACTIVITY}_I$(P1, P2).

4.5 Political Framing

Framing is a political strategy that describes the concept of how politicians word their statements in order to control the way the public views and discusses current issues. To investigate the intuition that the way politicians contextualize their tweets is strongly indicative of their stance on an issue, 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 with a political frame. As noted in Card et al. (2015), some tweets may fall into multiple frames. After all tweets are classified, we sum over the total number of each frame type and use the frames with the maximum and second largest counts as that politician's frames for that issue. The top two frames are used because for most 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).

4.6 Temporal Framing Patterns

While we expect politicians within a party to use similar frames per issue (as captured by the PSL predicate FRAME), it is also possible that politicians will use certain frames around events. For example, when a mass shooting occurs, we observe that Democrats will tweet about enacting gun legislation and typically frame these tweets as a matter of a needed preemptive action for public safety (which falls under the *Health and Safety* frame). In reaction to this, Republicans will tweet about protecting American citizens' rights to gun ownership, which falls under the *Constitutionality* frame. Therefore, we expect similarities and differences in framing around events to indicate similarities and differences in stances and agreement patterns. To capture this idea, we combine the approaches of Sections 4.4 and 4.5: we align the politicians' timelines per issue and compare the frames used to discuss the issue-related events. When two or more politicians use the same frame for an issue on the same day, we consider them to have similar temporal framing patterns. This is used as input to the PSL predicate $\text{SAMETEMPORALFRAME}_I(\text{P1, P2})$.

5 Global Models of Twitter Activity

Information obtained from the local models alone is not strong enough to quantify stance or agreement for politicians, as shown by our baseline measurements in Section 6. Therefore, we use PSL to build global connections among the output of the local models (which acts as weak supervision), resulting in global PSL models which successively build upon the previous model in order to obtain the highest accuracy possible. In addition to the PSL predicates representing the target output (PRO and SAMESTANCE_I) [4] and local models (as defined in Section 4), we also use directly observed information: party affiliation, denoted $\text{DEM}(\text{P1})$ for Democrat and $\neg\text{DEM}(\text{P1})$ for Republican, and $\text{SAMEPARTY}(\text{P1, P2})$ to denote if two politicians belong to the same political 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:

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

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 similarities between politicians based on temporal Twitter behaviors.

[4] 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.

5.2 Baseline: Using Local Classifiers Directly

To show that the local models do not provide enough information individually to make an accurate prediction, we implement a local baseline (LB) PSL model which does not take advantage of the global modeling framework. It instead learns weights over rules (shown in Table 3), which directly map the output of the local noisy classifiers described in Section 4 to PSL target predicates.

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 the Local Baseline Model.

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. The PSL rules listed in Table 4 are designed to capture this party based agreement. For some issues we initially assume Democrats (Dem) are for an issue, while Republicans (¬Dem) are against that issue, (e.g., ¬Dem(P1) →¬Pro(P1, 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 SameStance, 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: 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)
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)
Tweets(P1, Issue) ∧Tweets(P2, Issue) ∧Dem(P1) →Pro(P2, Issue)
Tweets(P1, Issue) ∧Tweets(P2, Issue) ∧¬Dem(P1) →¬Pro(P2,Issue)
TweetPos(P1, Issue) ∧TweetPos(P2, Issue) →SameStance$_I$(P1, P2)
TweetPos(P1, Issue) ∧TweetNeg(P2, Issue) →¬SameStance$_I$(P1, P2)
TweetPos(P1, Issue) ∧TweetPos(P2, Issue) ∧Dem(P1) →Pro(P2, Issue)
TweetNeg(P1,Issue) ∧TweetNeg(P2,Issue) ∧¬Dem(P1) →¬Pro(P2,Issue)
TweetPos(P1, Issue) ∧TweetPos(P2, Issue) ∧SameParty(P1, P2) →SameStance$_I$(P1, P2)
TweetPos(P1,Issue) ∧TweetNeg(P2,Issue) ∧¬SameParty(P1,P2) →¬SameStance$_I$(P1,P2)

Table 5: Subset of PSL Rules Used in Model 2.

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. The incorporation of this information allows Model 3 to overcome Model 2 inconsistencies between stance and sentiment (e.g., when someone is attacking their opposition). 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. FRAME indicates the frame used by that politician for different issues. Finally, SAMETEMPORALFRAME$_I$ conveys that two politicians use the same frames for an issue at the same time. More details on these predicates are in Sections 4.3, 4.4, 4.5, and 4.6 respectively.

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)
FRAME(P1, ISSUE) $\wedge$DEM(P1) $\rightarrow$PRO(P1, ISSUE)
FRAME(P1, ISSUE) $\wedge\neg$DEM(P1) $\rightarrow\neg$PRO(P1, ISSUE)
SAMETEMPORALFRAME$_I$(P1, P2) $\wedge$SAMEPARTY(P1, P2) $\rightarrow$SAMESTANCE$_I$(P1, P2)
SAMETEMPORALFRAME$_I$(P1, P2) $\wedge$PRO(P1, ISSUE) $\rightarrow$PRO(P2, ISSUE)

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

6 Experiments

6.1 Experimental Settings

Supervised Baseline: Previous works exploring stance classification typically predict stance based on an *individual item of text* (e.g., forum post or single tweet) in a *supervised* setting, making it difficult to directly compare to our approach. To facilitate comparison, we implemented a tweet-based supervised baseline, per issue. We labeled each tweet with the politician's stance (either for or against) on that tweet's issue. We trained an SVM on 80% of the politicians' tweets and tested on the remaining 20%, using 5-fold cross-validation. Because we aim to predict each politician's stance and *not* the stance of each tweet, we aggregated the SVM predictions by politician, i.e., the SVM predicts a stance for each tweet and the majority prediction among a politician's tweets is used as his or her stance. For agreement prediction, we compared this stance prediction across politicians to determine if the predicted stances agreed and whether or not this agreement was correct.

PSL Models: As described in Section 4, the data generated from the local models is used as weak supervision to initialize the PSL models described in Section 5. The Local Baseline model (LB) is initialized with only the information from the weak local models. We initialize Model 1 (M1), as described in Section 5.3, using knowledge of the politician's party affiliation. Model 2 (M2) builds upon (M1) by incorporating the results of the issue and sentiment analysis local models, as described in Sections 4.1 and 4.2 respectively. Model 3 (M3) combines all previous models with higher level knowledge of Twitter activity: tweet agreement (Section 4.3), temporal activity (Section 4.4), frames (Section 4.5), and temporal framing patterns (Section 4.6). We implement our PSL models to have an initial bias that candidates do not share a stance and are against an issue. Stances collected in Section 3.2 are used as the ground truth for evaluation of the predictions of the PSL models only, not for any form of supervision.

6.2 Experimental Results

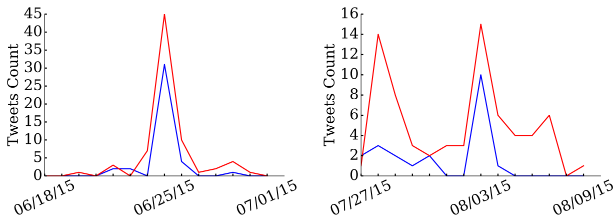
Results Per Issue: Table 7 presents the results of using the supervised baseline and our three proposed PSL models. While the supervised baseline results (SVM) are not directly comparable to our weakly supervised model, since the supervised model uses a different data split and approach, it does show that direct supervision does not lead to immediate prediction improvement and can result in weaker prediction scores. 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 11 of 16 issues; for prediction of agreement (SAMESTANCE$_I$), LB performs slightly lower overall, with only 9 of 16 issues predicted above chance. Using M1, we improve stance prediction accuracy for

Issue	STANCE (RESULTS OF PRO PREDICTION)					AGREEMENT (SAMESTANCE PREDICTION)				
	SVM	LB	M 1	M 2	M 3	SVM	LB	M 1	M 2	M 3
ABORTION	61.25	81.25	96.88	96.88	96.88	44.34	49.31	93.75	93.75	95.36
ACA	87.5	96.88	100	100	100	79.7	51.61	100	100	100
CONFEDERATE	16.56	34.38	78.12	84.38	87.5	0	51.31	69.6	77.7	80.18
DRUGS	48.13	87.5	78.12	88.88	96.88	44.34	50.42	63.6	84.07	84.07
ENVIRONMENT	69.06	53.12	78.12	78.13	81.08	65.86	45.16	65.59	68.75	71.37
GUNS	84.38	93.75	93.75	93.75	93.75	57.33	48.59	68.54	99.5	99.59
IMMIGRATION	73.44	37.5	81.25	81.25	86.36	51.82	53.62	68.55	69.06	69.56
IRAN	74.56	84.38	65.62	65.63	84.38	69.25	35.57	79.73	100	100
ISIS	80.0	40.32	76.28	93.75	94.44	74.19	59.68	76.28	76.28	90.04
MARRIAGE	33.44	62.5	90.62	90.62	90.9	12.5	50.57	87.12	87.13	87.43
NSA	21.25	37.5	53.12	53.12	61.54	2.61	34.15	49.2	56.66	60.08
PAY	34.38	84.38	84.38	89.47	90.62	29.59	64.30	72.92	74.31	80.31
RELIGION	42.81	75	68.75	81.25	81.25	56.89	47.62	76.24	76.46	79.44
SOCIAL SECURITY	35.31	28.12	78.12	78.13	78.13	0.91	53.76	73.25	90.03	90.88
STUDENT	0	93.75	96.88	96.88	96.88	0	51.61	100	100	100
TPP	0	62.5	62.5	62.5	62.5	0	45.43	48.39	54.64	65.32

Table 7: Stance and Agreement Accuracy by Issue. The SVM columns show the results of the tweet-based, supervised baseline. LB columns show the results when using only the weak local models. M1 columns are the results based on party line agreement, M2 columns are the results when adding Twitter activity to M1, and M3 columns are the results when adding higher level Twitter behaviors to M1 and M2.

	GLOBAL		REP		DEM	
	ST	AG	ST	AG	ST	AG
LB	68.36	52.49	66.80	49.10	69.92	44.86
M1	81.25	76.34	75.39	75.16	87.11	85.44
M2	85.16	87.30	81.25	84.26	89.06	91.37
M3	89.84	87.76	87.11	85.35	92.58	91.49

Table 8: Overall Accuracy for Stance (ST) and Agreement (AG) Prediction. GLOBAL represents the accuracy over all politicians, while REP and DEM refer to Republicans or Democrats only.



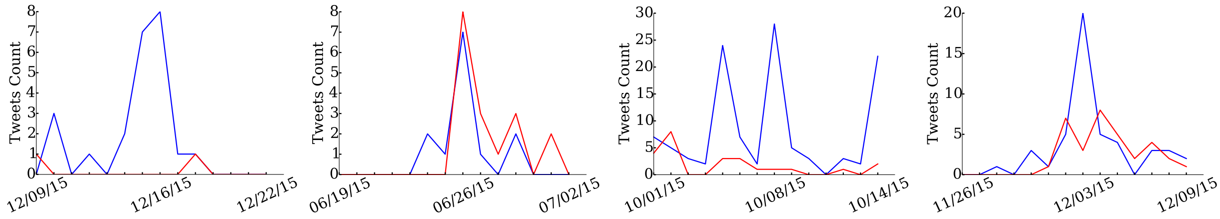
a ACA b Abortion

Figure 3: Temporal Twitter Activity by Party. The red and blue lines represent the temporal overlaps, or lack thereof, of Republicans and Democrats (respectively) in Twitter activity 1 week before and after a major event.

10 of the issues and agreement accuracy for all issues. M2 further improves the stance and agreement predictions for an additional 8 and 12 issues, respectively. M3, the combination of the previous models with Twitter behavioral features, increases the stance prediction accuracy of M2 for 9 issues and the agreement accuracy for 12 issues.

The final agreement predictions of M3 are notably improved over the initial LB for all issues, indicating that similarities and differences in Twitter behaviors help capture agreement and disagreement patterns among politicians. The final stance predictions of M3 are improved over all issues except Guns, Iran, and TPP. For Guns, the stance prediction remains the same throughout all models, meaning party information does not boost the initial predictions determined from Twitter based behaviors. For Iran, the addition of M1 and M2 lower the accuracy, but the temporal features from M3 are able to restore it to the original prediction. For TPP, this trend is likely due to the fact that all models incorporate 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. Even in cases where M1 and/or M2 remained steady or lowered the initial baseline result (e.g. stance for Religion and Pay), the final prediction by M3 is still higher than that of the baseline.

Overall Results: Table 8 presents our overall results for stance and agreement prediction in terms of accuracy. The Global score is the overall average for all politicians, while REP and DEM consider only Republicans or Democrats, respectively. Each model increases the accuracy of the previous model's prediction, showing that additional Twitter behavioral features can help overcome the strong party line biases captured by M1.



(a) Environment (b) Marriage (c) Guns (d) Guns

Figure 4: Temporal Twitter Activity by Party for Three Issues.

6.3 Effects of Framing and Temporal Activity Patterns

As shown in Table 7, performance for *some* issues does not improve in M3. 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 TPP, both parties share the *Economic* frame. In addition to similar framing overlap, the Twitter timeline for ACA also exhibits overlap, as shown in Figure 3(a). 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 3(b) 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 M3 to extract enough information to increase agreement prediction accuracy by 1.61%.

Figure 4(a) shows an example of one event for the Environment issue: when the mayor of Flint, Michigan declared a state of emergency over lead in the city's water supply. Due to different temporal patterns and frames for such events, the Environment accuracy improves across all models, as shown in Table 7. Similarly, Figure 4(b) shows the week before and after the Supreme Court ruled to uphold the legality of same-sex marriage. The two central peaks are shared by both parties, but each party also has one peak before (Democrats) or after (Republicans) the event. Additionally, both parties share the *Constitutionality* frame as their top frame, but the second most used frame is *Morality* for Republicans and *Fairness and Equality* for Democrats. These slight differences allow the M3 model to improve over the M2 prediction. Finally, Figure 4(c) shows the week before and after Democratic Senators pushed for gun control legislation after the Umpqua Community College shooting and Figure 4(d) shows tweets around the San Bernadino shooting. For these events, both parties exhibit different timeline patterns and frames. Consequently, these behavioral features dominate the stance prediction and allow agreement accuracy to reach 99.59%.

7 Conclusion

In this work we present a framework for modeling the dynamic nature of political discourse on Twitter. Though we focus on a small set of politicians and issues, our approach can be modified to handle additional politicians or issues, as well as those of other countries, by incorporating the proper 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. Contrary to previous works, which typically focus on a single aspect of this complex microblogging behavior, we build a holistic model connecting party line biases, temporal behaviors, and issue framing into a single predictive model which identifies fine-grained stances and agreement patterns. Despite having no direct supervision and using only intuitive local classifiers to bootstrap our global model, our approach results in a strong predictive model which helps shed light on political discourse within and across party lines.

Acknowledgments

We thank the anonymous reviewers for their thoughtful comments and suggestions.

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*.

Justin Grimmer. 2010. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. In *Political Analysis*.

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*.

Yoad Lewenberg, Yoram Bachrach, Lucas Bordeaux, and Pushmeet Kohli. 2016. Political dimensionality estimation using a probabilistic graphical model. In *Proc. of UAI*.

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*.

Axel Maireder and Julian Ausserhofer. 2013. National politics on twitter: Structures and topics of a networked public sphere. In *Information, Communication, and Society*.

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*.