Towards Tracking Political Sentiment through Microblog Data

Yu Wang
Emory University
yu.wang@emory.edu

Tom Clark
Emory University
tclark7@emory.edu

Jeffrey Staton
Emory University
jkstato@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

Abstract

People express and amplify political opinions in Microblogs such as Twitter, especially when major political decisions are made. Twitter provides a useful vehicle for capturing and tracking popular opinion on burning issues of the day. In this paper, we focus on tracking the changes in political sentiment related to the U.S. Supreme Court (SCOTUS) and its decisions, focusing on the key dimensions on support, emotional intensity, and polarity. Measuring changes in these sentiment dimensions could be useful for social and political scientists, policy makers, and the public. This preliminary work adapts existing sentiment analysis techniques to these new dimensions and the specifics of the corpus (Twitter). We illustrate the promise of our work with an important case study of tracking sentiment change building up to, and immediately following one recent landmark Supreme Court decision. This example illustrates how our work could help answer fundamental research questions in political science about the nature of Supreme Court power and its capacity to influence public discourse.

1 Background and Motivation

Political opinions are a popular topic in Microblogs. On June 26th, 2013, when the U.S. Supreme Court announced the decision on the unconstitutionality of the ”Defense of Marriage Act” (DOMA), there were millions of Tweets about the users' opinions of the decision. In their Tweets, people not only voice their opinions about the issues at stake, expressing different dimensions of sentiment, such as support or opposition to the decision, or anger or happiness. Thus, simply applying traditional sentiment analysis scales such as ”positive” vs. ”negative” classification would not be sufficient to understand the public reaction to political decisions.

Research on mass opinion and the Supreme Court is valuable as it could shed light on the fundamental and related normative concerns about the role of constitutional review in American governance, which emerge in a political system possessing democratic institutions at cross-purposes. One line of thought, beginning with Dahl (Dahl, 1957), suggests that the Supreme Court of the United States has a unique capacity among major institutions of American government to leverage its legitimacy in order to change mass opinion regarding salient policies. If the Dahl's hypothesis is correct, then the Supreme Court's same-sex marriage decisions should have resulted in a measurable change in opinion. A primary finding about implication of Dahl's hypothesis is that the Court is polarizing, creating more supportive opinions of the policies it reviews among those who supported the policy before the decision and more negative opinions among those who opposed the policy prior to the decision (Franklin and Kosaki, 1989) (Johnson and Martin, 1998).

We consider Twitter as important example of social expression of opinion. Recent studies of content on Twitter have revealed that 85% of Twitter content is related to spreading and commenting on headline news (Kwak et al., 2010); when users talk about commercial brands in their Tweets, about 20% of them have personal sentiment involved (Jansen et al., 2009). These statistical evidences imply that Twitter has became a portal for public to express opinions. In the context of politics, Twitter content, together with Twitter users'

Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media, pages 88–93,
Baltimore, Maryland USA, 27 June 2014. ©2014 Association for Computational Linguistics

information, such as user's profile and social network, have shown reasonable power of detecting user's political leaning (Conover et al., 2011) and predicting elections (Tumasjan et al., 2010). Although promising, the effectiveness of using Twitter content to measure public political opinions remains unclear. Several studies show limited correlation between sentiment on Twitter and political polls in elections (Mejova et al., 2013) (O'Connor et al., 2010). Our study mainly focuses on investigating sentiment on Twitter about U.S. Supreme Court decisions.

We propose more fine-grained dimensions for political sentiment analysis, such as supportiveness, emotional intensity and polarity, allowing political science researchers, policy makers, and the public to better comprehend the public reaction to major political issues of the day. As we describe below, these different dimensions of discourse on Twitter allows examination of the multiple ways in which discourse changes when the Supreme Court makes a decision on a given issue of public policy. Our dimensions also open the door to new avenues of theorizing about the nature of public discourse on policy debates.

Although general sentiment analysis has made significant advances over the last decade (Pang et al., 2002) (Pang and Lee, 2008) (Liu, 2012) (Wilson et al., 2009), and with the focus on certain aspects, such as intensity (Wilson et al., 2004), irony detection (Carvalho et al., 2009) and sarcasm detection (Davidov et al., 2010), analyzing Microblog content such as Twitter remains a challenging research topic (Reyes et al., 2012) (Vanin et al., 2013) (Agarwal et al., 2011). Unlike previous work, we introduce and focus on sentiment dimensions particularly important for political analysis of Microblog text, and extend and adapt classification techniques accordingly. To make the data and sentiment analysis results accessible for researchers in other domain, we build a website to visualize the sentiment dynamics over time and let users download the data. Users could also define their own topics of interest and perform deeper analysis with keyword filtering and geolocation filtering.

We present a case study in which our results might be used to answer core questions in political science about the nature of Supreme Court influence on public opinion. Political scientists have long been concerned with whether and how Supreme Court decisions affect public opinion and discourse about political topics (Hoekstra, 2003) (Johnson and Martin, 1998) (Gibson et al., 2003). Survey research on the subject has been limited in two ways. Survey analysis, including panel designs, rely on estimates near but never on the date of particular decisions. In addition, all survey-based research relies on estimates derived from an instrument designed to elicit sentiment – survey responses, useful as they are, do not reflect well how public opinion is naturally expressed. Our analysis allows for the examination of public opinion as it is naturally expressed and in a way that is precisely connected to the timing of decisions.

Next, we state the problem more formally, and outline our approach and implementation.

2 Problem Statement and Approach

2.1 Political Sentiment Classification

We propose three refinements to sentiment analysis to quantify political opinions. Specifically, we pose the following dimensions as particularly important for politics:

- Support: Whether a Tweet is ***Opposed***, ***Neutral***, or ***Supportive*** regarding the topic.
- Emotional Intensity: Whether a Tweet is emotionally ***Intense*** or ***Dispassionate***.
- Sentiment Polarity: Whether a Tweet's tone is ***Angry***, ***Neutral***, or ***Pleased***.

2.2 Approach

In this work, each of the proposed measures is treated as a supervised classification problem. We use multi-class classification algorithms to model Support and Sentiment Polarity, and binary classification for Emotional Intensity and Sarcasm. Section 3.2 describes the labels used to train the supervised classification models. Notice some classes are more interesting than the others. For example, the trends or ratio of opposed vs. supportive Microblogs are more informative than the factual ones. Particularly, we pay more attention to the classes of *opposed*, *supportive*, *intense*, *angry*, and *pleased*.

2.3 Classifier Feature Groups

To classify the Microblog message into the classes of interest, we develop 6 groups of features:
Popularity: Number of times the message has been

posted or favored by users. As for a Tweet, this feature means number of Retweets and favorites.
Capitalization and Punctuation.
N-gram of text: Unigram, bigram, and trigram of the message text.
Sentiment score: The maximum, minimum, average and sum of sentiment score of terms and each Part-of-Speech tags in the message text.
Counter factuality and temporal compression dictionary: This feature counts the number of times such words appear in the message text.
Political dictionary: Number of times a political-related word appears in the message text.
We compute sentiment scores based on SentiWordNet[1], a sentiment dictionary constructed on WordNet.[2] Political dictionary is built upon political-related words in WordNet. As in this paper, we construct a political dictionary with 56 words and phrases, such as "liberal", "conservative", and "freedom" etc.

3 Case Study: DOMA

Our goal is to build and test classifiers that can distinguish political content between classes of interest. Particularly, we focus on classifying Tweets related to one of the most popular political topics, "Defence of Marriage Act" or DOMA, as the target. The techniques can be easily generalized to other political issues in Twitter.

3.1 Dataset

In order to obtain relevant Tweets, we use Twitter streaming API to track representative keywords which include "DOMA", "gay marriage", "Prop8", etc. We track all matched Tweets generated from June 16th to June 29th, immediately prior and subsequent to the DOMA decision, which results in more than 40 thousand Tweets per day on average.

3.2 Human Judgments

With more than 0.5 million potential DOMA relevant Tweets collected, we randomly sampled 100 Tweets per day from June 16th to June 29th, and 1,400 Tweets were selected in total. Three research assistants were trained and they showed high agreement on assigning labels of relevance, support, emotional intensity, and sentiment polarity after training. Each Tweet in our samples was labeled by all three annotators. After the labeling, we first removed "irrelevant" Tweets (if the Tweet was assigned "irrelevant" label by at least one annotator), and then the tweets with no major agreement among annotators on any of the sentiment dimensions were removed. As a result, 1,151 tweets with what we consider to be reliable labels remained in our dataset (which we expect to share with the research community).

3.2.1 Annotator Agreement

The Fleiss' Kappa agreement for each scale is reported in Table 1 and shows that labelers have an almost perfect agreement on relevance. Support, emotional intensity, and sentiment polarity, show either moderate or almost perfect agreement.

Measure	Fleiss' Kappa
Relevance	0.93
Support	0.84
Intensity	0.54
Polarity	0.49

Table 1: Agreement (Fleiss' Kappa) of Human Labels.

3.3 Classification Performance Results

We reproduce the same feature types as previous work and develop the political dictionary feature for this particular task. We experimented with a variety of automated classification algorithms, and for this preliminary experiment report the performance of Naïve Bayes algorithm (simple, fast, and shown to be surprisingly robust to classification tasks with sparse and noisy training data). 10-fold cross validation are performed to test the generalizability of the classifiers. Table 2 reports the average precision, recall and accuracy for all measures. Sarcasm is challenging to detect in part due to the lack of positive instances. One goal in this study is to build a model that captures trends among the different classes. In Section 3.4, we will show that the trends of different measures estimated by the trained classifier align with the human annotated ones over time.

3.4 Visualizing Sentiment Before and After DOMA

One natural application of the automated political sentiment analysis proposed in this paper is tracking public sentiment around landmark U.S. Supreme Court decisions. To provide a more reliable estimate, we apply our trained classifier on all relevant Tweets in our collection. More than

[1] http://sentiwordnet.isti.cnr.it/

[2] http://wordnet.princeton.edu/

Value	Prec. (%)	Rec. (%)	Accuracy(%)
Supportive (48%)	73	74	68
Neutral (45%)	76	67	
Opposed (7%)	17	30	
Intense (31%)	56	60	73
Dispassionate (69%)	81	79	
Pleased (10%)	48	31	69
Neutral (79%)	84	78	
Angry (11%)	24	45	

Table 2: Performance of Classifiers on Each Class.

2.5 million Tweets are estimated in four proposed measures. Figure 1 shows the distribution of on-topic Tweet count over time. The Supreme Court decision triggered a huge wave of Tweets, and the volume went down quickly since then.

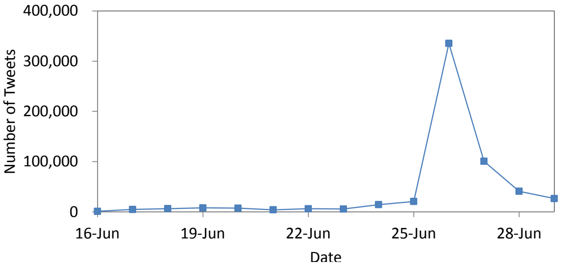

Figure 1: Number of “Gay Marriage” Tweets Over Time.

Figures 2 and 3 visualize both the human labeled trends and the ones obtained by the classifier for the classes “Supportive” and “Intense”. In both figures, the peaks in the predicted labels generally align with the human-judged ones. We can see the supportiveness and intensity are both relatively high before the decision, and then they decline gradually after the Supreme Court decision.

Figure 3 shows the volume of intensive Tweets detected by our trained model has a burst on June 22rd, which is not captured by human labeled data. To investigate this, we manually checked all Tweets estimated as “intensive” on June 22rd. It turns out most of the Tweets are indeed intensive. The reason of the burst is that one Tweet was heavily retweeted on that day. We do not disclose the actual tweet due to its offensive content.

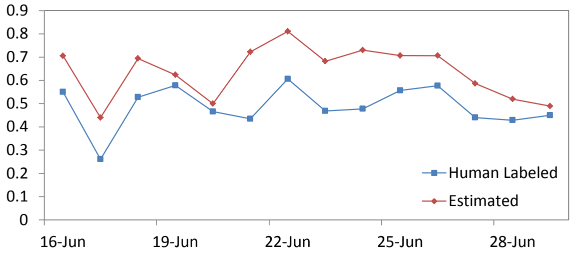

Figure 2: Percentage of “Supportive” Tweets Over Time.

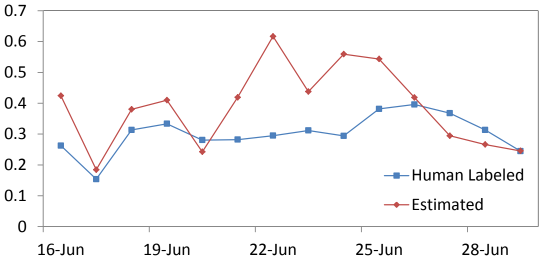

Figure 3: Percentage of “Intense” Tweets Over Time.

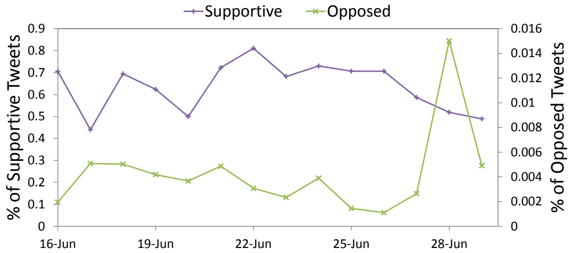

Figure 4: Comparison between “Supportive” and “Opposed” Trends.

Figure 4 plots the trends of “supportive” and “opposed” Tweets in different scales. According to the Supreme Court decision, the “supportive” group wins the debate. Interestingly, instead of responding immediately, the “loser” group react and start Tweeting 2 days after the decision. These trends indicate that “winner” and “loser” in the debate react differently in time and intensity dimensions.

We believe that our estimates of sentiment can be used in various ways by political scientists. The “positivity bias” (Gibson and Caldeira, 2009) model of Supreme Court opinion suggests that the Court can move public opinion in the direction of its decisions. Our results possibly indicate the opposite, the “polarizing” model suggested by (Franklin and Kosaki, 1989) and (Johnson and Martin, 1998), where more negative opinions are observed after the decision (in Figure 4), at least for a short period. By learning and visualize political sentiments, we could crystalize the nature of the decision that influences the degree to which the Supreme Court can move opinion in the direction of its decisions.

4 An Open Platform for Sharing and Analyzing Political Sentiments

Figure 5 shows a website[3] that visualizes political sentiments over time. The website shows several popular U.S. Supreme Court cases, such as “gay marriage”, “voting right act”, “tax cases”,

[3] http://www.courtometer.com

etc., and general topics, such as “Supreme Court” and “Justices”. Each of the topics is represented by a list of keywords developed by political science experts. The keywords are also used to track relevant Tweets through Twitter streaming API. To let users go deeper in analyzing public opinions, the website provides two types of real-time filtering: keywords and location of Tweet authors. After applying filters, a subset of matched Tweets are generated as subtopics and their sentiments are visualized. The example filtering in Figure 5 shows the process of creating subtopic “voting right act” out of a general topic “Supreme Court” by using keyword “VRA”. We can see that the volume of negative Tweets of “voting right act” is higher than the positive ones, compared to the overall sentiment of the general Supreme Court topic. Once an interesting subtopic is found, users can download the corresponding data and share with other users.

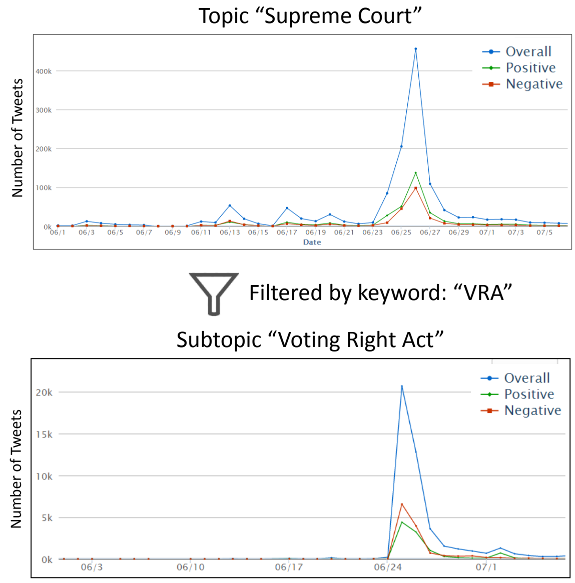

Figure 5: We build a website that visualizes political sentiments over time and let users create “subtopics” by using keyword and location filters.

5 Conclusions

In this paper we considered the problem of political sentiment analysis. We refined the notion of sentiment, as applicable to the political domain, and explored the features needed to perform automated classification to these dimensions, on a real corpus of tweets about one U.S. Supreme Court case. We showed that our existing classifier can already be useful for exploratory political analysis, by comparing the predicted sentiment trends to those derived from manual human judgments, and then applying the classifier on a large sample of tweets – with the results providing additional evidence for an important model of Supreme Court opinion formation from political science.

This work provides an important step towards robust sentiment analysis in the political domain, and the data collected in our study is expected to serve as a stepping stone for subsequent exploration. In the future, we plan to refine and improve the classification performance by exploring additional features, in particular in the latent topic space, and experimenting with other political science topics.

ACKNOWLEDGMENTS The work of Yu Wang and Eugene Agichtein was supported in part by DARPA grants N11AP20012 and D11AP00269, and by the Google Research Award.

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. *Sentiment Analysis of Twitter Data.* In Proceedings of the Workshop on Language in Social Media (LSM).

Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. 2009. *Clues for detecting irony in user-generated contents: oh...!! it’s ”so easy” ;-).* In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion.

M.D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer. 2011. *Predicting the Political Alignment of Twitter Users* In Proceedings of IEEE third international conference on social computing

Robert Dahl. 1957. *Decision-Making in a Democracy: The Supreme Court as National Policy-Maker.* Journal of Public Law.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. *Semi-supervised Recognition of Sarcastic Sentences in Twitter and Amazon.* In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL).

Charles H. Franklin, and Liane C. Kosaki. 1989. *Republican Schoolmaster: The U.S. Supreme Court, Public Opinion, and Abortion.* The American Political Science Review.

James L Gibson, and Gregory A Caldeira. 2009. *Citizens, courts, and confirmations: Positivity theory and the judgments of the American people.* Princeton University Press.

James L Gibson, Gregory A Caldeira, and Lester Kenyatta Spence. 2003. *Measuring Attitudes toward the*

United States Supreme Court. American Journal of Political Science.

Valerie Hoekstra. 2003. *Public Reaction to Supreme Court Decisions*. Cambridge University Press.

Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. *Micro-blogging As Online Word of Mouth Branding*. in CHI '09 Extended Abstracts on Human Factors in Computing Systems.

Timothy R. Johnson, and Andrew D. Martin. 1998. *The Public's Conditional Response to Supreme Court Decisions*. American Political Science Review 92(2):299-309.

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. *What is Twitter, a Social Network or a News Media?*. in Proceedings of the 19th International Conference on World Wide Web (WWW).

Yu-Ru Lin, Drew Margolin, Brian Keegan, and David Lazer. 2013. *Voices of Victory: A Computational Focus Group Framework for Tracking Opinion Shift in Real Time*. In Proceedings of International World Wide Web Conference (WWW).

Bing Liu. 2012. *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies.

Yelena Mejova, Padmini Srinivasan, and Bob Boynton. 2013. *GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media*. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM).

B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. 2010. *From tweets to polls: Linking text sentiment to public opinion time series*. In Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM).

Bo Pang, and Lillian Lee. 2008. *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. *Thumbs up? sentiment classification using machine learning techniques*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).

Antonio Reyes, Paolo Rosso, and Tony Veale. 2012. *A multidimensional approach for detecting irony in Twitter*. Language Resources and Evaluation.

Swapna Somasundaran, Galileo Namata, Lise Getoor, and Janyce Wiebe. 2009. *Opinion Graphs for Polarity and Discourse Classification*. TextGraphs-4: Graph-based Methods for Natural Language Processing.

Aline A. Vanin, Larissa A. Freitas, Re-nata Vieira, and Marco Bochernitsan. 2013. *Some clues on irony detection in tweets*. In Proceedings of International World Wide Web Conference (WWW).

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. *Recognizing Contextual Polarity: an exploration of features for phrase-level sentiment analysis*. Computational Linguistics.

Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2004. *Just how mad are you? Finding strong and weak opinion clauses*. In Proceedings of Conference on Artificial Intelligence (AAAI).

Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM).