

インフォーマルな政治的意見テキスト分類のためのセンチメント分析の方法の対価

トニー・マレン      ロバート・マルフ

津田塾大学      San Diego State University

インターネットの知的環境の特徴として、インフォーマルな政治的談話はますます重要度を増している。そのやりとりから意味のある情報を自動的に抜き出すことは、依然困難な課題である。我々が得た結果によると、投稿自体の意見内容よりも投稿者間の修辞学上の関係が、分類のためには有益である。本論文では、インフォーマルの政治的領域の感情分析を用い、書き手の政治的志向によってテキストを分類する目的で進行中のいくつかの実験について解説する。有名な感情分析方法論に基づくアプローチを使用した実験および、これらを活用する動機、起こりうる障害を記述する。

A consideration of sentiment analysis methodologies for classification of political opinion texts

Tony Mullen      Robert Malouf

Tsuda College, Tokyo      San Diego State University

Informal political discourse has become an ever more important feature of the intellectual landscape of the Internet, automatic extraction of meaningful information from such exchanges remains a formidable challenge. Past results we have achieved indicate that rhetorical relationships between posters are much more salient and useful for classification purposes than the actual opinion content of the posts themselves. In this paper we describe our ongoing experiments in classifying texts by the political orientation of the writers by means of sentiment analysis for the informal political domain. Experiments with several approaches based on well-known sentiment analysis methodologies are described, along with the motivations for applying them in this domain and a description of the possible difficulties we anticipate.

## Introduction

Our research has been motivated by the increasingly prominent role played by informal political discourse in the intellectual landscape of the Internet. Until recently, much of the text available on the WWW was either professionally edited or followed the conventions of edited text. Newspapers, magazines, corporate and government publications, and academic papers, and even personal and hobby sites produced by amateurs follow fairly rigid standards for formatting, style, and orthography. Furthermore, they are intended to be read by a large anonymous audience which shares only the most general public context. This makes them comparatively easy to process with relatively little specific background knowledge, both for human readers who may be referred to the site by a search engine, and for automated methods such as question answering or text mining systems.

Now, a large portion of American political discourse has begun to take place in environments such as discussion forums, social networking sites, and chat rooms, where

content is submitted by users. Websites such as MySpace and YouTube, both of which are driven entirely by user-submitted content, have already begun to play a key role in the 2008 presidential campaign itself.

Unlike edited text, language use in informal web texts is typically conversational, is often non-standard or idiosyncratic, and is highly contextualized, depending on rich background of shared knowledge and assumptions. Furthermore, communicative skills, linguistic and otherwise, vary much more drastically among informal posters than they do between professional reporters and editorial writers. These facts have shown to be significant in terms of what kind of information can be used to serve as a basis for automatic classification of political orientations of texts and users.

While some work has been done on sentiment analysis for political texts (Efron, 2004; Efron, 2003), the extent to which this task differs from more conventional sentiment analysis tasks, and the pertinence of conventional approaches to the current task have not been exhaustively investigated. We have been interested in exploring the also possibilities of exploiting sentiment values of phrases and clauses, taking cues from methods such as those presented in Wilson, et al.(2005), Nasukawa (2003), and Turney (2005), although the task we would like to perform is considerably more complex than that of simple binary favorability assessment. In this paper we follow work reported in (Malouf & Mullen, 2007) using a data set of political discourse data from an online American politics discussion group.

## **Analysis of politically relevant sentiment**

There are many applications for recognizing politically-oriented sentiment in texts. These applications include analyzing political trends within the context of a given natural language domain as a means of augmenting opinion polling data; classifying individual texts and users in order to target advertising and communications such as notices, donation requests or petitions; and identifying political bias in texts, particularly in news texts or other purportedly unbiased texts. This last use is particularly pertinent to evaluating the reliability of information sources, since it is widely assumed that an excess of political bias is a corrupting factor on the reliability of an information source.

Many of the challenges of the present task are analogous, but not always identical, to those faced by traditional sentiment analysis. It is well-known that people express their feelings and opinions in oblique ways. Furthermore, unlike opinion as addressed in conventional sentiment analysis, which focuses on favorability measurements toward specific entities, political attitudes generally encompass a collection of favorability judgments towards a variety of different entities and issues.

## **Challenges in processing the data**

The data we analyze has two distinct defining characteristics: its predominantly political content and its informality. Each of these qualities introduces challenges and methods of

addressing these challenges can sometimes interfere with each other. One of the difficulties with analysis of informal text is dealing with the considerable problem of rampant spelling errors. This problem is compounded when the work is in a domain such as politics, where jargon, names, and other non-dictionary words are standard. The domain of "informal politics" introduces jargon all of its own, incorporating terms of abuse, pointed respellings, The difficulties of analysis on the word level percolate to the level of part-of-speech tagging and upwards, making any linguistic analysis challenging.

There is another possible, more subtle level on which this verbal "informality" may be indicative of deeper difficulties to be surmounted. Conventional sentiment analysis approaches rely on a coherent and consistent presentation of opinions. The work of Turney (2002) indicated that sentiment analysis tasks that are less emotionally charged, such as reviews electronic products, are easier to handle than more emotional, impressionistic tasks such as classifying movie reviews.

### **Political sentiment analysis as a classification task**

For the present task, we conducted tests using several classification schemes. We used both the hand-modified self-descriptions as they stood, and we used a more general classification of `textsl{right, left,}` and `textsl{other}`, which was composed of people who described themselves as "centrist", "libertarian" or "independent." The hand-modification we did on the self-descriptions was usually straightforward, although in one instance a self-described "Conservative Democrat" was modified to "conservative." If there had been enough conservative Democrats in the data to justify it, this classification probably should have been allowed to stand as a distinct self-described class, and generalized to the `textsl{other}` class.

### **Data resources**

We created a database of political discourse downloaded from [www.politics.com](http://www.politics.com). The database consists of approximately 77,854 posts organized into topic threads, chronologically ordered, and identified according to author, author's stated political affiliation. Furthermore, the posts are broken down into smaller chunks of text based on typographical cues such as new lines, quotes, boldface, and italics, which represent segments of text which may be quotes from other authors. Each text chunk of three words or greater is identified as quoted text or non-quoted text based upon whether it is identical to a substring in a previous post by another poster. The database contains 229,482 individual text chunks, about 10 percent of which (22,391 chunks) are quotes from other posts.

The total number of individual posters is 408. The number of posts by each author follows an inverse power-law distribution, with 77 posters (19%) logging only a single post. The greatest number of posts logged by a single poster is 6,885 posts, followed by the second greatest number of posts at 3,801 posts.

RIGHT 34%	Republican	53
	Conservative	30
	R-fringe	5
LEFT 37%	Democrat	62
	Liberal	28
	L-fringe	6
OTHER 28%	Centrist	7
	Independent	33
	Libertarian	22
	Green	11
	Unknown	151

Fig. 1. Distribution of posts by the general class and by a slightly modified version of the writers' own self-descriptions.

In addition to the main dataset used for training and testing, additional data from the web was used to support spelling-correction. For this, we used 6481 politically oriented syndicated columns published online on right and left leaning websites `www.townhall.com` and `www.workingforchange.com` (4496 articles and 1985 articles, respectively). We also used a wordlist of email, chat and text message slang, including such terms as "lol," meaning "laugh out loud."

### Text classification

To test the effectiveness of standard text classification methods for predicting political affiliation, we divided the users into the two general classes RIGHT (Republican, conservative, and r-fringe) and LEFT (Democrat, liberal, and l-fringe), setting aside the centrist, independent, green, and libertarian users. We then used the naive Bayes text classifier Rainbow (McCallum, 1996), to predict the political affiliation of a user based on the user's posts. There were 96 users in the LEFT category and 89 in the RIGHT, so a baseline classifier which assigned the category LEFT to every user would yield 52.17% accuracy. The NB text classifier gave an accuracy of 63.59% a modest (though statistically significant) improvement over the baseline. Mullen & Malouf (2007) established that improvements in modeling can be obtained by means of clustering groups of users based upon patterns of agreement and disagreement with each other. We would like to deepen this analysis by looking more closely at the texts themselves and determine whether it is feasible to determine political orientation using sentiment analysis methods.

### Sentiment analysis

In our efforts to employ sentiment analysis tools on informal political natural language data, we have made a number of assumptions. Some of these assumptions may not be warranted, and part of our research should consist of trying to highlight and explore the degree to which the assumptions are valid. The basic assumption we make is that political orientations are, in some way, closely related to opinions in the familiar sense of the word. We do not necessarily assume that political views themselves are truly binary; it may be that they are the results of a number of favorability judgments about various topics and entities, which are in turn analogous to the kinds of opinions traditionally dealt with in sentiment analysis. The other assumption we make, which is perhaps even more open to question, is that informal political writing like the texts we analyze is in some way written in an analogous way to traditional opinion texts. That is, we assume that people posting to political discussions are doing so in order to express their opinions. This would seem at first blush reasonable to assume, but some of our results so far have caused us to consider the possibility that this assumption may not be warranted.

### **Liberal/conservative values**

Our first approach to using Sentiment Analysis, described in Malouf & Mullen (2007) attempted to classify users using an approach inspired by the PMI-IR method used in Turney (2002) method. Rather than using the method to find positive and negative semantic orientations for phrases, the approach used an analogous approach to look for "conservative" and "liberal" semantic orientations. The semantic orientations derived for phrases were largely accurate, but they did not prove to be useful to classification of users or texts containing them.

### **Multiple binary-valued judgments**

We are currently investigating an approach to sentiment analysis of political opinions, in which we consider each user to be a vector of real-valued PMI-IR-derived sentiment values for 110 hand-selected divisive political topics. These vectors are then used as training data for a Support Vector Machine model.

A value for the "semantic orientation" (SO) of each phrase is identified by finding occurrences of target phrase near one of a fixed set of positive words such as "excellent" and a fixed set of negative words such as "poor" in a very large reference corpus (we used the Waterloo MultiText system, designed expressly for this purpose). Pointwise mutual information (PMI) of each phrase with positive and negative words is calculated. The PMI of the phrase with negative words and the PMI of the phrase with positive words are combined to create an overall *semantic orientation* (SO) of the phrase, which is a positive number for "positive" or "good" words and a negative number for "negative" or "bad" words". The overall orientation of a text is considered to be the average of the SOs of the phrases in the text. Turney's method yielded overall accuracy of 74.39%, although the results varied widely

across domains (60%--85%).

We create a list of political "hot topics" by hand. These included words and phrases we judged to be divisive along right/left lines in American politics, such as "Iraq war", "abortion", "death penalty", "gun control", "Bush", "Cheney", "Clinton", "Kennedy" etc. In total, 110 word and phrase sets (synonyms, plurals, shortenings, etc, were considered together) were collected. For each user, a "representative text" was created for each hot topic, made up of sentences written by that user that included a reference to the hot topic. Each of these representative texts was analyzed using the standard PMI-IR method, yielding a positive or negative number which we took to represent the user's opinion on that issue. These numbers are then used as feature values in an SVM model.

As of the present writing, we have not had success with this approach, and we are currently investigating the reasons why this approach may fail to live up to what appears to be its potential. As Turney (2002) pointed out, reviews such as movie reviews, which involve a considerable amount of emotion and subjectivity, and which are usually written with considerable rhetorical flourish, are not well modeled by sentiment analysis methods. The preponderance of identifiable terms of abuse and slang terms in this data may be indicative of heavy use of other kinds of rhetorical styles such as sarcasm, which present profound difficulties for pattern-based sentiment analysis.

We hope that a way can be found of combining values in an SVM that will help to sift out some of the noise and bring focus onto meaningful patterns, but it is possible also that by bringing in too many features that did not help in modeling, we may have further obscured what useful information may have been hidden in some of the features.

## Conclusions

Our results suggest that information gained from the discourse relations between posters is of use in identifying the political sentiment of the posts. More interestingly, it suggests that the antagonisms and that arise between posters may be more explicitly identifiable than the underlying beliefs that inform their arguments. Although not especially surprising given the complexity of the task, it is notable how resistant informal political discourse has been to sentiment analysis methods. Of course, these results are far from conclusive, and there are many reasons why the sentiment analysis approach described here may not work. Also, counter to our previously stated assumptions, it may be the case that the rhetorical goals of political debators are not as analogous to those of reviewers as we had assumed. Indeed, in many cases posters are not writing to inform others of their opinions; after a few exchanges posters often treat their actual opinions as given and proceed to defend their views or attack other views, both of which are fundamentally different rhetorical goals from the goal of expressing, announcing, or describing an opinion in the first place. We could expect this to result in considerable obfuscation of the actual opinions in the written text.

In any case, it appears to be clear that some of the necessary conditions for good sentiment analysis--that opinions be clearly and reasonably consistently presented, with as

little misleading rhetoric as possible—are not sufficiently met by the present data. In order to test our first assumption, that political views are analogous to other kinds of opinions, it may be necessary to attempt to apply these sentiment analysis techniques to a different kind of textual data, which more closely resembles traditional opinion text data from a rhetorical standpoint.

This work is supported by KAKENHI 18700158, Grant-in-Aid for Young Scientists (B) provided by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

## References

- Efron, M.: Cultural orientation: Classifying subjective documents by cocitation analysis. In: AAI Fall Symposium on Style and Meaning in Language, Art, and Music (2004)
- Efron, M., Zhang, J., Marchionini, G.: Implications of the recursive representation problem for automatic concept identification in on-line government information. In: Proceedings of the ASIST SIG-CR Workshop (2003)
- Malouf, R., Mullen, T.: Graph-based user classification for informal online political discourse. In: Proceedings of the 1st Workshop on Information Credibility on the Web, Miyazaki, Japan (2007)
- McCallum, A.K.: Bow: A toolkit for statistical language modeling, text retrieval, classification, and clustering, <http://www.cs.cmu.edu/~mccallum/bow> (1996)
- Nasukawa, T., Yi, J.: Sentiment analysis: Capturing favorability using natural language processing. In: The Second International Conferences on Knowledge Capture (K-Cap 2003). (2003)
- Turney, P.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, ACL (2002)
- Turney, P.: Measuring semantic similarity by latent relational analysis. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05), Edinburgh, Scotland (2005)
- Wilson, T., Wiebe, J., Hoffman, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT-EMNLP. (2005)