A Strife of Interests: Challenges for Sentiment Analysis of Informal Political Discourse

Tony Mullen
Department of Computer Science
Tsuda College, Japan
mullen@tsuda.ac.jp

Robert Malouf
Department of Linguistics and Oriental Languages
San Diego State University
rmalouf@mail.sdsu.edu

Introduction

- Motivation and challenges for political sentiment analysis
- The politics.com dataset
- Classification by agreement grouping (graph-based)
- Sentiment-based approaches
- Discussion and conclusions

Sentiment analysis

- Identifying opinions, favorability judgments, etc. in NL texts.
- Usually applied to texts with explicit opinion intent: reviews, customer feedback, etc.
- Treated as a classification task
- Supervised, unsupervised ML, shallow linguistics.

Analyzing political opinion

- Possible applications:
  - Analyzing political trends/Augmenting opinion polling data
  - Targeting advertising and communications such as notices, donation requests, or petitions
  - Identifying political bias, e.g. in news texts

What is informal political discourse?

- Informal political discourse can be found in
  - Online forums
  - Newsgroups/mailing lists
  - Blogs
  - Online publications reader feedback sections
  - Social Networking Services
- Often organized as linear threads by topic.
- Discourse is “informal”; written quickly, as thought
- Overall discourse not real-time, but individual exchanges may be near real-time.

Idiosyncrasies of informal political discourse

- Informal
  - Rampant spelling errors
  - Casual usage (sentence fragments, etc.)
- Political
  - Jargon, names, non-dictionary terms
- Informal *and* political
  - Specific jargon, terms of abuse (“wingnuts”, “moonbats”)
  - Satirical re-spellings of known words (“Raygun”, Repugnicans”, “Dumbocrats”)

Sentiment analysis of informal political discourse

- What is “political opinion?”
  - SA often considers a binary “thumbs up” vs “thumbs down” classification
  - This is too simple to represent political opinion.
- Political attitudes encompass a variety of favorability judgments
- Relations between judgments are not always clear; e.g., in the US political domain anti-abortion judgment often corresponds to pro-death penalty judgment.

Possible goals for political SA

- Aside from binary judgments about a specific issue, candidate, or proposal, we might want to:
  - Identify political party affiliation
  - Classify according to some more general taxonomy, e.g. right vs left
  - Gauge the “extremeness” or distance from a politically centrist position of the writer’s views
  - Evaluate the degree of confidence with which the writer expresses views
  - Evaluate the degree of agreeability/argumentativeness with which the writer communicates
  - Identify particular issues of special importance to the writer

Classifying political attitudes

- As a preliminary task, we opted for the simplest classification scheme we could think of:
  - right
  - left
  - other
- Many viewpoints do not fit tidily on the left/right line, and “other” is so general as to be essentially noise

The data

- Data from the (now defunct) www.politics.com discussion site
- 77,854 posts organized by topic thread
- 408 individual posters
- Number of posts follows a Zipf-like distribution, with 19% of posters logging only a single post.
- Greatest number of posts by a single poster is 6885, second is 3801

Identifying quotes

- Each post broken into “chunks” based upon typographical cues such as new lines, quotes, boldface, and italics, to identify sections of the post which are quoted from previous posts.
- Chunks of three words or greater which are complete substrings of previous posts are considered quotes.
- The database is broken into 229,482 individual chunks, of which 22,391 are identified as quotes from other posts.

Supplementary data

- Additional data from the web was used to support spelling correction
  - 6481 politically oriented syndicated columns from right and left leaning websites, to provide professionally edited spellings of domain specific terms
  - A wordlist of email, chat, and text message slang, including such terms as “lol” meaning “laugh out loud”

Political affiliation in the data

- Posters have a self-described political affiliation.
- After some hand-editing, nine modified labels were identified:
  - Republican
  - Conservative
  - R-fringe
  - Democrat
  - Liberal
  - L-fringe
  - Centrist
  - Independent
  - Libertarian

Classes to stated affiliation

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
Unknown			151

Naïve Bayes lexical model

- First, we used naïve Bayes to classify posts lexically as Left or Right
- “Other” users were disregarded
- Total number of users were 96 left, and 89 right, so the baseline was 51.9%
- Lexical model performed at 60.4%

Observations on the lexical model

- Unlike with topic identification, arguments from both sides of an issue use many of the same terms.
- Irregular spellings are harmful to lexical models, necessitating far more training data.
- Skewed distribution of posting frequency means that frequent posters are better modeled than infrequent posters

Some adjustments

- Restricting experiments to frequent posters (500+ words)
  - Baseline 50%
  - Naïve Bayes: 61.38%
  - With spelling correction: 64.48%
- Human gold standard 87.5% for all users, 91% for frequent posters

Quote patterns

- Of 41,605 posts 4,583 contained quoted material
- Strong tendency to quote users from opposite end of political spectrum
  - Left quoted right: 62.2%
  - Right quoted left: 77.5%

Classification by quote

- For frequent posters:
  - For those who quote/are quoted: 83.53
  - Overall: 79.38
- However, this assumes that we know the class of the quoted poster

Using user citation graph information

- Created a graph with each user as a node and each quote an edge
- Singular value decomposition on graph's adjacency matrix to compute a "citation space" in which distances between users could be measured
- Derived equivalence classes via alliance/agreement patterns

Using user citation graph information

- Graph-based clustering + NB yielded 68.48% accuracy for all users, 73% for frequent posters

Sentiment Analysis of Texts

- Assumptions
  - Political attitudes are (the same as|analogous to|composed of) the kind of opinions found in reviews
  - Political discussion is rhetorically similar in some significant respect to opinion/review writing

PMI-IR Sentiment Analysis

- Turney (2002) used PMI-IR to do sentiment analysis for reviews
- Unsupervised approach to finding positive or negative nuance for words and phrases (Semantic Orientation)
- PMI-IR values found between phrase and “excellent” and “poor”
- Turney used entire Internet (via AltaVista) as training data

PMI-IR Sentiment Analysis

$$\mathrm{PMI}(word_1, word_2) = \log_2\left(\frac{\mathrm{p}(word_1 \,\&\, word_2)}{\mathrm{p}(word_1)\mathrm{p}(word_2)}\right)$$

$$\begin{aligned}\mathrm{SO}(phrase) &= \mathrm{PMI}(phrase, \text{"excellent"}) \\ &\quad - \mathrm{PMI}(phrase, \text{"poor"})\end{aligned}$$

Simple PMI-IR inspired political classifcation

$$\mathrm{SO}(\mathit{phrase}) = \mathrm{PMI}(\mathit{phrase}, \text{"liberal"}) - \mathrm{PMI}(\mathit{phrase}, \text{"conservative"})$$

- Derived SO values from Reuters corpus
- Results considerably below baseline

Simple PMI-IR inspired political classifcation

- Possible reasons for poor performance
  - Wrong choice of contrast terms?
  - Inadequate training data?
  - Deeper assumptions mistaken?

Single-Issue PMI-IR Vectors

- Assume that political attitudes are collections of positive/negative judgements on single, hot-button issues
- Draw up a list of politically contentious words/terms/names
- From each poster, select all sentences containing each of these terms
- Evaluate using PMI-IR to get an SO score for each concept
- SVM model with resulting feature vectors

Single-Issue PMI-IR Vectors

- Created approximately 100 contentious concepts by hand, intuitively likely to distinguish right from left in American political discussion.
- Turney’s Waterloo Multitext system to derive SO values
- No deviation from the baseline

What are the problems?

- As usual, data is sparse
- Political opinions expressed more obliquely than, e.g. movie reviews?
- Rhetorical goals different?
  - Reviews are written to express/describe/justify opinions
  - Political discussion posts treat underlying opinions as given and focus on convincing and/or attacking

Conclusions

- Patterns of agreement/disagreement more salient than actual opinion content
- Political discussion more than just a description of opinions on various topics
- PMI-IR based methods not promising for informal political text analysis

Thank you