Predicting Persuasiveness in Political Discourses

Carlo Strapparava, Marco Guerini, Oliviero Stock

FBK-irst, 38050, Povo, Trento, Italy

{strappa, guerini, stock}@fbk.eu

Abstract

In political speeches, the audience tends to react or resonate to signals of persuasive communication, including an expected theme, a name or an expression. Automatically predicting the impact of such discourses is a challenging task. In fact nowadays, with the huge amount of textual material that flows on the Web (news, discourses, blogs, etc.), it can be useful to have a measure for testing the persuasiveness of what we retrieve or possibly of what we want to publish on Web. In this paper we exploit a corpus of political discourses collected from various Web sources, tagged with audience reactions, such as applause, as indicators of persuasive expressions. In particular, we use this data set in a machine learning framework to explore the possibility of classifying the transcript of political discourses, according to their persuasive power, predicting the sentences that possibly trigger applause. We also explore differences between Democratic and Republican speeches, experiment the resulting classifiers in grading some of the discourses in the Obama-McCain presidential campaign available on the Web.

1. Introduction

Persuasive Natural Language Processing focuses on the use of language for inducing desired beliefs and behaviors (e.g. approval, agreement, appreciation) in the receivers. Indeed, human communication has often the purpose of persuading or convincing. For example, politicians search for peoples approval, or advertising copywriters for influencing people purchases. Nowadays, with the huge amount of textual material that flows on the Web (news, discourses, blogs, etc.), it can be useful to have a measure for testing the persuasiveness of what we retrieve or possibly of what we want to publish on Web.

In this paper we approach the task of automatically predicting the impact of political discourses. In particular, in political speeches, the audience tends to react or resonate to signals of persuasive communication, including an expected theme, a name or an expression. In public speaking, the communication, if well-planned and practiced, can be a memorable and pleasurable event for both the speaker and the audience. Even if nonverbal communication plays a role, nonetheless language and words are what the speaker uses to convey the core message to the audience. Likewise as public speakers understand the impact of their own words using audience feedback, we exploit CORPS, a freely available corpus (Guerini et al., 2008), which contains political speeches tagged tagged with audience reactions, such as applause, as indicators of persuasive expressions. We use this data set in a machine learning framework to explore the possibility of classifying the transcript of political discourses, according to their persuasive power, predicting the sentences that possibly trigger applause. We also explore differences between Democratic and Republican speeches, experiment the resulting classifiers in grading some of the discourses in the Obama-McCain presidential campaign available on the Web.

2. Persuasion and NLP

Some works on persuasion and NLP have mainly focussed on Natural Language Generation. Persuasive text generation deals with the production of texts that are meant to affect the behavior of the receiver (Reiter et al., 2003).

Opinion mining is a topic at the crossroads of information retrieval and computational linguistics concerned with the identification of opinions (either positive or negative) expressed in a document (Wilson et al., 2004; Wiebe and Mihalcea, 2006; Breck et al., 2007; Pang and Lee, 2008). While opinion mining deals with texts that are meant to persuade, nonetheless its focus is on polarity (valence) recognition for evaluative language retrieval.

While there is vast theoretical research on politicians' rhetorics, only recently there has been a growing interest in bridging the gap between qualitative analysis of political communication and computational linguistics (Cousins and Mcintosh, 2005; Bligh et al., 2004).

The automatic analysis of political communication is mainly focused on *text categorization*. Text categorization deals with the task of assigning a document to a pre-defined set of categories, such as determining party position in a text (e.g. Republicans or Democrats), see for example the work by (Purpura and Hillard, 2006; Jiang and Argamon, 2008).

3. A Corpus of Political Speeches

For the experiments in this paper we exploit CORPS (CORpus of tagged Political Speeches), a resource freely available for research purposes (Guerini et al., 2008), which contains political speeches tagged with audience reactions (e.g. applause, standing-ovation, booing). The collected texts come from various Web sources (e.g. politicians' official sites, News web sites). The corpus was built relying on the hypothesis that tags about public reaction, such as APPLAUSE, are indicators of hot-spots where persuasion attempts succeeded or, at least, a persuasive attempt had been recognized by the audience.

Given that the corpus is composed of transcriptions of speeches mostly given at public mass gatherings, in general the audience is favorable to the speakers and the context is one of support. Of course, by giving value to the

Tag
APPLAUSE
SPONTANEOUS-DEMONSTRATION
STANDING-OVATION
SUSTAINED-APPLAUSE
CHEERS
BOOING

Table 1: List of the main tags

Figure 2: Temporal distribution of the speeches

audience reactions, we do not mean that the audience is actually effectively persuaded of some ideas or induced to do something that it did not believe in beforehand, even if the audience can be reassured, inspired or helped in making sense of events. To the contrary, the audience tends just to react to signals, including an expected theme, a name, an expression, the tone of the voice. Often the signals are creative, in the sense that the speaker may have produced new forms through creative rhetorical elaboration, but eventually they are recognized. Therefore the audience, so to say, resonates to a fragment of speech, which is meant to be of a persuasive genre. So we believe that there is a wealth of material that, by virtue of the validation provided by the audience reaction, can be used by a machine to automatically learn a model of persuasive language. This can be exploited to effectively persuade somebody, or simply to reproduce politicians' speech or be used for analyzing the pragmatic characteristics of a novel text.

At present, there are about 900 speeches in the corpus and about 2.2 millions words (see Figure 1 for a survey on main speakers' number of speeches and Figure 2 for the time distribution of the speeches). The speeches are all in native English language, and all represent monological situations (i.e. there is only one speaker addressing the audience).
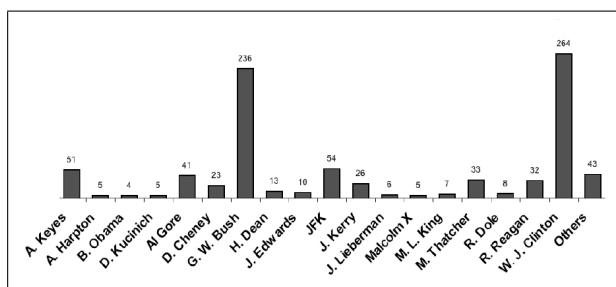

Figure 1: Number of speeches per speaker

The corpus proved to be a helpful resource for qualitatively analyzing political persuasive communication. See (Guerini et al., 2008) for more details.

4.   Experiments

We conducted a series of experiments on the corpus in a machine learning framework to explore the feasibility of (i) predicting the passages in the discourses that trigger a positive audience reactions, (ii) distinguishing in the corpus Democrats from Republicans, (iii) checking the audience
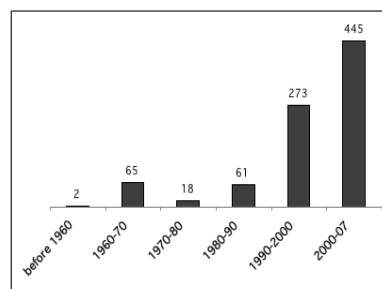
reaction classification if training is made on adverse party speeches (e.g. training on Republican speeches and testing on Democratic ones), and (iv) experimenting the classifiers on plain and typical non-persuasive texts taken from British National Corpus and on discourses from the last Obama-McCain presidential campaign.

For all the experiments we used Support Vector Machines (SVM) framework, in particular SVM-light under its default settings (Joachims, 1998).

Data set preprocessing. To reduce data sparseness, we pos-tagged all the corpus (using TnT pos-tagger[1]). So we considered lemmata instead of tokens in the format *lemma#POS*. In these experiments, we included all the tokens, i.e. we did not make any frequency cutoff or feature selection. Then we divided all the speeches into fragments of about four sentences[2]. The obtained chunks are then labeled as Neutral (i.e. no tag), and Positive-ironical (i.e. a tag that groups all positive audience reaction: APPLAUSE, STANDING-OVATION, SUSTAINED-APPLAUSE, CHEERS, SPONTANEOUS-DEMONSTRATION, LAUGHTER)[3].

Finally we got a total of 37,480 four-sentences chunks, roughly equally partitioned into the two considered labels. This accounts for a baseline of 0.50. In all the experiments we randomly split the corpus in 80% training and 20% test.

4.1.   Experiments on CORPS

Democrats vs. Republicans. First we simply tested the separation between Democratic and Republican speeches. This experiment was mainly conducted to see if the SVM setting, used for the next experiments, suitably distinguishes between the two parties, given that the topics dealt by the speakers are often quite similar. The corpus containing a total of 18,384 chunks coming from Republican speeches and 19,096 Democratic ones. From Table 2 we see that four-sentences chunks are enough to detect Republicans vs. Democrats distinction with a performance of 0.804 (F1 measure).

Positive audience reaction. Then we tested the capability

---

[1]http://www.coli.uni-saarland.de/~thorsten/tnt

[2]The chunks are about four sentences long, because if a tag is present in the fragment the chunk ends at that point.

[3]In the experiments we did not consider the Negative-focus tags (e.g. Booing), since there is only a small amount of them.

	Precision	Recall	F1
Democrats	0.842	0.756	0.797
Republicans	0.773	0.854	0.811
micro	0.804	0.804	0.804

Table 2: Republicans vs. Democrats (4-sentences chunks)

of predicting a positive audience reaction. As explained above in this case the tags to be classified are `Neutral` and `Positive-ironical`. First we experimented on all the corpus (Table 3), then we split the corpus in two: Democrats and Republicans. So we verified the classification on the two different parts separately (Tables 4 and 5) and in addition the case of cross-classification (i.e. training on Democrats and testing on Republicans and vice versa - Tables 6 and 7). In all the cases we randomly split in 80/20 training-test partition.

We see that the persuasive impact of speeches are quite general and, as shown in the cross-classification results, to a certain degree independent from the party of the speakers.

	Precision	Recall	F1
Positive-Ironical	0.646	0.683	0.664
Neutral	0.676	0.641	0.658
micro	0.660	0.660	0.660

Table 3: Positive-Ironical vs. Neutral (4-sentences chunks - Repub/Democ corpus)

	Precision	Recall	F1
Positive-Ironical	0.660	0.766	0.709
Neutral	0.663	0.549	0.601
micro	0.661	0.661	0.661

Table 4: Positive-Ironical vs. Neutral (4-sentences chunks - Republicans Only)

	Precision	Recall	F1
Positive-Ironical	0.666	0.674	0.670
Neutral	0.686	0.680	0.683
micro	0.676	0.676	0.676

Table 5: Positive-Ironical vs. Neutral (4-sentences chunks - Democrats Only)

4.2. Exploiting the Classifier

Testing on non-persuasive texts. In order to test the capabilities of distinguishing persuasive from non-persuasive texts, we conducted some experiments running the classifier, trained on CORPS, on about 7300 four-sentences chunks extracted from typical non-persuasive texts of the British

	Precision	Recall	F1
Positive-Ironical	0.642	0.632	0.637
Neutral	0.579	0.599	0.589
micro	0.612	0.612	0.612

Table 6: Positive-Ironical vs. Neutral (4-sentences chunks - Training on Democrats, Test on Republicans)

	Precision	Recall	F1
Positive-Ironical	0.625	0.660	0.642
Neutral	0.658	0.626	0.641
micro	0.641	0.641	0.641

Table 7: Positive-Ironical vs. Neutral (4-sentences chunks - Training on Republicans, Test on Democrats)

Total chunks	7243
Positive-Ironical	784
Neutral	6459
Prec/Rec/F1	0.892

Table 8: Classification on BNC

	Obama	McCain
Positive-Ironical	2372	2360
Neutral	68	80
Total chunks	2440	2440

Table 9: Classification on Obama/McCain last campaign speeches

National Corpus[4], so considered labeled as `Neutral`. Table 8 summarizes the results.

Obama/McCain presidential campaign. As a last experiment, we could not refrain from testing the classifier trained on CORPS for Obama's and McCain's speeches taken from the 2008 presidential campaign[5]. These speeches were not labeled (i.e. it was not possible to train on that political campaign), so the experiment should be regarded as generic test. The speeches were divided into four sentence chunks similarly to other data sets. The results show that the persuasive content of the speeches was quite high, with slightly better results for Obama.

5. Conclusions

In this paper, we explored the applicability of computational approaches to the recognition of persuasive language.

---

[4]We extract the chunks from A00 to A0H texts of BNC sources.

[5]We considered the discourses from the official candidates web sites: http://www.barackobama.com/speeches/index.php, http://www.johnmccain.com/informing/news/speeches/

Specifically, we investigated whether automatic classification techniques represent a viable approach for predicting the impact of a text, in particular for distinguish between persuasive and neutral texts.

To this purpose we exploited a corpus of political speeches, as examples of long and elaborated persuasive texts. The discourses are tagged with audience reactions (e.g. applause, standing-ovation, booing). Then we conducted a series of experiments for predicting the passages in the discourses that trigger a positive audience reactions. The results show that this could be a viable approach for studying the persuasive power of discourses.

The list of themes for future work includes for example: extracting specific persuasive lexicon; temporal analysis on how persuasiveness varies before and after key events (e.g. audience appreciation of Obama's/McCain's discourses after the financial crisis); and including some rhetorical cues in the presented framework. These techniques can allow us to develop systems for predicting how a new discourse will be evaluated by a given audience, and for suggesting how to modify it to increase its impact.

6.   References

M. C. Bligh, J. C. Kohles, and J. R. Meindl. 2004. Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly*, 15(2):211–239.

E. Breck, Y. Choi, and C. Cardie. 2007. Identifying expressions of opinion in context. In *Proceedings of IJCAI-2007*, Hyderabad, India, January.

K. Cousins and W. Mcintosh. 2005. More than typewriters, more than adding machines: Integrating information technology into political research. *Quality and Quantity*, 39:581–614.

M. Guerini, C. Strapparava, and O. Stock. 2008. CORPS: A corpus of tagged political speeches for persuasive communication processing. *Journal of Information Technology & Politics*, 5(1):19–32.

M. Jiang and S. Argamon. 2008. Preliminary semantic analysis of political blogs. In *Proceedings of ICWSM-08*, Seattle, March/April.

T. Joachims. 1998. Text categorization with Support Vector Machines: learning with many relevant features. In *Proceedings of the European Conference on Machine Learning*.

B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

S. Purpura and D. Hillard. 2006. Automated classification of congressional legislation. In *Proceedings of the Seventh International Conference on Digital Government Research*, San Diego, CA.

E. Reiter, S. Sripada, and R. Robertson. 2003. Acquiring correct knowledge for natural language generation. *Journal of Artificial Intelligence Research*, 18(491–516).

J. Wiebe and R. Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of 21st International Conference on Computational Linguistics (ACL-2006)*.

T. Wilson, J. Wiebe, and R. Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of AAAI*, pages 761–769.