Tracking Sentiment Evolution on User-Generated Content: A Case Study on the Brazilian Political Scene

Diego Tumitan, Karin Becker

Instituto de Informatica - Universidade Federal do Rio Grande do Sul, Brazil
{dctumitan, karin.becker}@inf.ufrgs.br

Abstract. Opinion mining, which aims the automatic processing of subjective information, has become a key field. In the political context, the awareness of people's sentiment towards their representatives, public organizations, political parties or politicians, can support political decisions, campaign moves, government policies, marketing strategies, etc. This paper describes a case study over opinions expressed about politicians as a reaction to news. Our ultimate goal was to detect whether user comments on an on-line newspapers reflect external indicators of public acceptance (e.g. vote intention). The paper outlines the approach used to identify and classify sentiment in news comments written in Portuguese language and to correlate it to external indicators, and discusses the main results for this case study.

Categories and Subject Descriptors: I.7 [**Document and Text Processing**]: Miscellaneous

Keywords: news, opinion mining, sentiment analysis, user-generated content

1. INTRODUCTION

Web users are no longer mere consumers of information. They interact with other users, expressing opinions and sentiment about various entities, such as products, brands, political figures, etc. This rich content can influence others and therefore, opinion mining, which aims at automatically processing subjective content, has become a very active field [Tsytsarau and Palpanas 2012].

Early work on opinion mining concentrated on products/services reviews [Tsytsarau and Palpanas 2012]. More recent works focus on specific entities (e.g. politicians, brands) on social networks [Pak and Paroubek 2010; Guerra et al. 2011] and news [Godbole et al. 2007]. The overall goal is to capture and track the general sentiment over the time, as represented by some metric, towards a target entity supporting many types of application.

In the political context, the awareness of people's opinion about their representatives, public organizations, political parties or politicians, can support political decisions, campaign moves, government policies, marketing strategies, etc. Traditional approaches involve expensive (and therefore infrequent) polls for detecting politicians' popularity, government approval, vote intention, etc. The potential of opinion mining for more up-to-date and broad opinion perspective has been demonstrated with regard to social medias such as Twitter or Facebook, and many commercial tools are available (e.g. TweetSentiments[1], Sentimonitor[2]). Less attention has been paid to user-generated content over news.

This paper describes a case study over opinions expressed in Portuguese about politicians as a reaction to news. Our ultimate goal was to detect whether user comments on on-line news reflects

[1]http://twittersentiment.appspot.com
[2]http://www.sentimonitor.com

external indicators of public acceptance (e.g. vote intention). We analyzed data referring to the 2012 mayoral elections of São Paulo, expressed as comments on a major on-line newspaper (Folha on-line), and used the external indicators provided by Datafolha polls. Our findings for this specific case study were: a) people do not tend to comment about the specific news content, but rather express their feelings in general about politics, politicians and their parties; b) there was an overall frustration over the current state of affairs, with a majority of negative comments; c) unlike other medias (e.g. Twitter, Facebook), very few people use this media support candidates, and d) considering the metrics developed, the sentiment has moderate correlation with vote intention for the candidates ellected for the second round. This case study is part of an on-going research about mechanisms for detecting and predicting sentiment evolution, based on user-generated comments written in Portuguese language.

The remainder of this paper is organized as follows. Section 2 contains related work. We outline the adopted approach in Section 3, and describe the case study in Section 4. Conclusions and future work are addressed in Section 5.

2. RELATED WORK

Opinion mining involves detecting subjective content, classifying its polarity, and summarizing the overall sentiment. Polarity classification relies on dictionary-based, machine-learning or statistical methods [Tsytsarau and Palpanas 2012]. The former is the most common one, but its results are dependent on the quality of sentiment lexicons. Classification can be at document or sentence-level. The latter is appropriate when a same document express opinions on several entities.

Many works have addressed the identification of sentiment about entities. Sentiment expressed in news towards an specific entity is analyzed and tracked in the system discussed in [Godbole et al. 2007]. The approach is based on an expansion technique over WordNet [Fellbaum 2010], which is not available for Portuguese. User-generated content on tweets are addressed in [Pak and Paroubek 2010; Narr et al. ; Guerra et al. 2011]. The former two propose a language-independent machine-learning approach, but which requires a training corpus. To eliminate that need, a transfer-knowledge approach is proposed in [Guerra et al. 2011], in which the sentiment is derived from the social relations between known pro/against opinion holders. On the political context, a study analyzes the results of elections in Germany with regard to the emotion expressed in tweets with mentions to political parties and candidates [Tumasjan et al. 2010]. The approach uses a linguistic tool that is not available for the Portuguese language.

One of the few works addressing user-generated content related to news in Portuguese language is [Sarmento et al. 2009], in which the authors create a set of lexico-syntactic patterns to identify the polarity of sentences. All used sentences are from comments related to a political newspaper, but the authors' goal is to create a reference corpus for political opinion mining.

Our work differs from the previous ones in that we address comments expressed as a reaction to news, in Portuguese language, and verify whether they correlate to external indicators of public acceptance. Towards this end, we develop a case study using the mayoral election of 2012.

3. APPROACH OUTLINE

In this paper, we describe a case study that tracks the general public sentiment towards political figures over the time, based on the perception of comments extracted from news regarding the Brazilian political scene. Our ultimate goal was to detect whether user comments on on-line newspapers reflect and correlate with external indicators of public acceptance (e.g. vote intention). This is a preliminary result of a broader on-going research, in which techniques for forecasting changes of attitude are under investigation. Figure 1 shows an overview of the proposed analysis approach. Notice that the process is highly iterative, in which returns to previous steps are necessary to improve results.

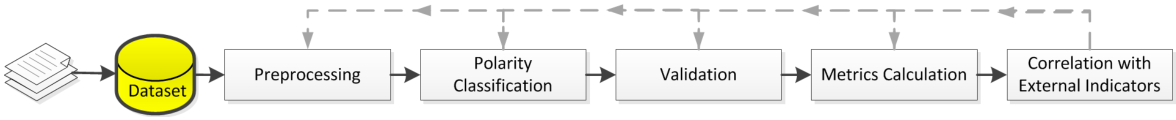

Fig. 1. Overview of the iterative proposed analysis approach.

Preprocessing: considering a dataset composed of news and their respective user comments, this step involves tasks to improve or discard user-generated comment, such as removing duplicated or excessively short comments, identification of slangs or domain-specific vocabulary, unification of all names and aliases for observed candidates (e.g nicknames, mean expressions), identification of misspelled or disguised words (e.g. cursing), among others.

Polarity Classification: encompasses breaking comments into sentences, identifying the ones with mentions to candidates, and classifying sentences' polarity. Sentence-level classification is necessary for most comments that involve more than one entity. We used a Portugal Portuguese sentiment lexicon, enriched with Brazilian and domain-specific terms. Sentiment words are identified, summarized (positive terms are added, and negative terms are subtracted), and the resulting sentiment is assigned to the target. Each target is thus associated with a set of positive, negative and neutral sentences.

Validation: In the lack of an annotated corpus, and considering the extent of the content to be analyzed, we adopted a sample-based validation. We randomly selected a set of sentences, and 3 different people annotated them using the same set of instructions. Only sentences with at least 2 agreements were used to validate.

Metrics Calculation: the polarized sentences are finally summarized to compose different metrics. We developed and experimented with different metrics, displayed in the first two columns of Table I. A suitable metric should reflect the general public sentiment, and a possible way to evaluate the best metric is through their correlation of with external indicators of public acceptance.

Correlation with External Indicators: Each domain may have different indicators that express overall sentiment, and which can be used to assess the results. In this case study, we adopted a typical election indicators such as voting intention and rejection rate. Job approval, popularity, economical indexes are other possible examples.

4. CASE STUDY

4.1 Dataset and Data Preprocessing

Dataset. We composed the dataset with news and comments extracted from the on-line version of Folha de São Paulo, one of the main newspapers in Brazil. We extracted news on the Brazilian Mayoral Elections of São Paulo, related mostly with the candidates, their parties, campaign moves, etc. We used Google Reader[3] as the indexer of Folha.com's political news section, called *Poder* (Power). These news cover the period from September 1st, to October 7th, 2012, which corresponds to the first round of the election. We extracted 583 news and 36,108 respective user's comments. We only considered the three leading candidates: Celso Russomanno, Fernando Haddad and José Serra [Datafolha 2012].

Data Preprocessing. we removed 3,808 duplicated comments (detected using the Cosine Similarity index) and 7,185 unreasonably short comments (less than 3 words or empty). After manual inspection, we realized other issues to be handled, such as transformation of words that were disguised by special characters (e.g. c@n@lh@ - scoundrel); misspelling or bad use of accentuation; use of regional (e.g. "petralha", "malufista") and idiomatic expressions ("é o cara" for the expression "he's the man") denoting sentiment. We also identified variations on candidate mentions (e.g. Serra, Zehserra),

[3]http://google.com/reader

Table I. Proposed metrics and correlation between sentiment metrics and vote intention/rejection rate.

Description	**Formula**	**Haddad**	**Serra**	**Russomanno**
Ratio of positive sentiment of an entity to the negative sentiment of the same entity	$s_d = \frac{pos_e}{neg_e}$ (1)	0.57/0.44	0.54/-0,14	0.12/-0.04
Ratio of positive sentiment to the total sentiment	$s_d = \frac{pos_e}{pos_e + neg_e}$ (2)	0.56/0.40	0.54/-0.16	0.08/-0.04
Ratio of negative sentiment to the total sentiment	$s_d = \frac{neg_e}{pos_e + neg_e}$ (3)	-0.56/-0.40	-0.54/0.16	-0.08/0.04
Ratio of positive sentiment of an entity to the positive sentiment of all entities	$s_d = \frac{pos_e}{pos_{entities}}$ (4)	0.09/0.07	0.22/0.29	0.39/-0.29
Ratio of negative sentiment of an entity to the negative sentiment of all entities	$s_d = \frac{neg_e}{neg_{entities}}$ (5)	-0.35/-0.23	0.16/0.34	0.16/-0.04

some of them with implied sentiment (e.g. Vampiserra, Malhaddad, as mean aliases), which were handled using regular expressions based on the candidates' names. Domain-specific sentiment vocabulary was added to the used lexicon along the process. Lastly, comments were broken into 79,752 sentences.

4.2 Target Identification and Sentiment Classification

Each sentence containing a mention to a candidate and sentiment words (9,758 sentences) was then polarized. We adopted SentiLex-PT [Silva et al. 2012], which contains 7,014 lemmas e 32,347 inflected forms for Portugal Portuguese. Each entry has a polarity (1, -1 and 0). To improve our results, over the time we added regional and domain-dependent terms to this dictionary. We tried different approaches for handling negations (e.g. "not good"), but no experiment yielded good results yet.

4.2.1 *Approach Validation.* To validate the sentence classification performance, we randomly selected 600 sentences that contained mentions to the candidates. The annotators were three graduate students majored in computer science, with no previous experience on corpus annotation. They were instructed to base their classification on what was explicitly written, disregarding any assumption about political entities or parties [Sarmento et al. 2009], so that their political background would not interfere in their judgment. The resulting gold-standard contained 482 sentences classified as negative, 72 as positive, and 46 as neutral. We discarded 3% of the sentences due to the lack of agreement.

4.2.2 *Performance Assessment.* Considering our gold-standard, we developed different experiments to classify the sentiment of the sentences, including attempts to handle negation adverbs (e.g. not good, never excelled). Only the best 3 results are discussed here due to space limitations. Results are summarized in Table II, which does not display the results for the neutral class.

In the *"Baseline"* experiment, we straightforwardly applied the co-occurrence method, with no significant terms preprocessing. We obtained fairly good results in terms of precision for the negative sentences, but we were not satisfied neither with the precision of the positive sentences, nor with the recall for both positive and negative classes. The next two experiments report two attempts developed for improving positive sentences classification and recall.

In the *"Modified Lexicon"* experiment, we manually analyzed the top 1,000 more frequent words that were not in the SentiLex-PT. As a result, we selected and added to the lexicon 268 new words and idiomatic expressions. Our results for the positive class significantly improved.

The *"Without Accentuation"* experiment adopts the refined lexicon and addresses users' accentuation typos. We removed all the accentuation from both comments and sentiment lexicon. We significantly improved the recall for negative sentences, while maintaining the same precision for both classes. Thus,

Table II. Experiments results according to Accuracy (P), Micro-Averaged F1 (Mi-A), Macro-Average F1 (Ma-A), Precision (P), Recall (R) and F-score (F).

Variation	A(%)	Mi-A(%)	Ma-A(%)	Polarity	P (%)	R (%)	F(%)
Baseline	35.54	46.17	39.19	Positive	17.35	21.79	19.32
				Negative	89.22	37.76	53.06
Modified Lexicon	43.21	50.05	46.17	Positive	26.02	**65.38**	**37.23**
				Negative	**90.52**	39.63	55.12
W/O Accentuation	**52.14**	**58.52**	**51.02**	Positive	**26.99**	56.41	36.51
				Negative	90.18	**51.45**	**65.52**

according to Accuracy, Micro-averaged F1 and Macro-Averaged F1, the results of this experiment constitute a significant improvement with regard to all previous attempts. This is thus the approach used for tracking sentiment in the remaining of this paper.

4.2.3 *Polarity Classification Error Analysis.* In general, our method did not perform well in classifying positive sentences. User-generated content is full of typographic errors, and thus the lexicon may contain the sentiment word, but the term is not recognized. We minimize this issue at some extent by disregarding accentuation. We also observed the importance of sentiment words related to an specific context [Godbole et al. 2007], with many terms specific to the city of São Paulo.

An important issue is that many sentences express comparative opinions (e.g. *"X is better than the candidate Y, because he is less involved in corruption"*, where the negative polarity of *"corruption"* will be related to both candidates). Finally, irony and sarcasm is a hard problem.

4.3 Sentiment Tracking

Considering the 9,758 polarized sentences resulting from the previous steps, we calculated the value for each metric of Table I per candidate and per day. Finally, we examined whether they correlated over the time with election external indicators (vote intention and rejection rate), as provided by poll results provided by Datafolha [Datafolha 2012] using Pearson correlation.

In order to visually analyze the trend and compare their evolution more easily, we applied z-score for both data since they are at different scales. We also distributed the polls values linearly due to the different data granularity (eight polls exist in the considered period - x axis of Figure 2). This choice represents the assumption that no change occurred in public opinion in between the poll results' publications, which may not be correct. Figure 2 shows the overlap of the sentiment ratio (dashed line) and vote intention (solid line) for the two candidates elected for the second round. The peak of sentiments observed on Sept. 25^{th} and Oct. 3^{rd} correspond to comments about news on the tie between Serra and Haddad. The correlation of each metric for both vote intention and rejection rate for each candidate is displayed in the last three columns of Table I. Considering the vote intention, the first three metrics worked fairly well for both Serra and Haddad, with moderate correlation. This means that vote intention increases along with positive comments and decrease of negative comments. However, no metric presented a consistent result for Russomanno. Actually, we observed that people just quited commenting about him near the election day, a fact that may have influenced these results. As for rejection rate, we observed almost no correlation, meaning that either these are not good metrics, or rejection rates reveal intrinsic feelings that are harder to influence with comments on news.

5. CONCLUSION AND FUTURE WORK

This paper described a case study over opinions expressed about politicians as a reaction to news. Our ultimate goal was to detect whether user comments on an on-line newspapers reflect external indicators of public acceptance. We presented the method used, the best results of the experiments

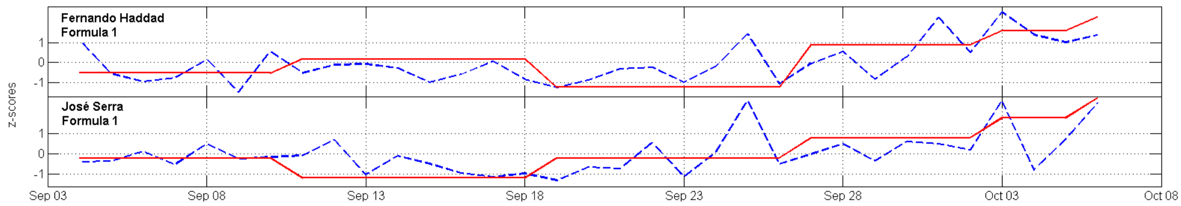

Fig. 2. Overlap of positive sentiment ratio (dashed) and vote intention (solid), normalized using z-score

developed, and the results for the correlation according to several metrics, which has shown a moderate correlation with vote intention for the candidates elected for the second round.

Our primarily expectation was that user-generated comments would reveal the authors' opinion about the respective news post. We realized that most comments refers to frustration about politics in general; transference of opinion by the candidate's association to other corrupt politicians or parties; political scandals; poor previous administration; etc.

We also expected that authors would support or make opposition to candidates. Support was very rare, and the authors would rather debate which candidates were less worst. Thus, this media seems to present a different role and impact if compared to Twitter or other social networks [Tumasjan et al. 2010; Pak and Paroubek 2010; Guerra et al. 2011].

Another challenge was the presence of sentiment words and idiomatic expression that are exclusive to the Brazilian language, political context and even a city. Reactions to Russomanno were very different from the ones towards the other candidates, due to his association to religion, a situation which was unique to the city of São Paulo. Replicating this experiment to other cities or elections involves the hard task of identifying contextual, regional and domain-dependent terms. The process of acquiring domain-specific vocabulary is laborious, and subject to errors. New approaches need to be considered for improving the classification results.

This work is part of an on-going broader research on deriving models that detects changing patterns on the attitude towards a subject. To overcome limitations of the present study, we need to consider other cities and candidates, as well as other on-line news and social medias, thus extending the type of comments addressed and their scope. We are currently experimenting techniques to develop a predictive model for public acceptance of political figures based on the sentiment expressed.

REFERENCES

Datafolha. Intenção de voto para prefeito de são paulo, 2012. Retrieved November 19, 2012.

Fellbaum, C. *WordNet.* Springer, 2010.

Godbole, N., Srinivasaiah, M., and Skiena, S. Large-scale sentiment analysis for news and blogs. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*. Vol. 2, 2007.

Guerra, P. H. C., Veloso, A., Jr., W. M., and Almeida, V. From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In *Proceedings of the KDD*. pp. 150–158, 2011.

Narr, S., Hülfenhaus, M., and Albayrak, S. Language-independent twitter sentiment analysis.

Pak, A. and Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, 2010.

Sarmento, L., Carvalho, P., Silva, M., and de Oliveira, E. Automatic creation of a reference corpus for political opinion mining in user-generated content. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM, pp. 29–36, 2009.

Silva, M., Carvalho, P., and Sarmento, L. Building a sentiment lexicon for social judgement mining. *Computational Processing of the Portuguese Language*, 2012.

Tsytsarau, M. and Palpanas, T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 2012.

Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International aaai conference on weblogs and social media*. pp. 178–185, 2010.