

# Analyzing Political Sentiment on Twitter

**Martin Ringsquandl and Dušan Petković**

University of Applied Sciences Rosenheim  
 petkovic@fh-rosenheim.de  
 martin.ringsquandl@googlemail.com

## Abstract

Due to the vast amount of user-generated content in the emerging Web 2.0, there is a growing need for computational processing of sentiment analysis in documents. Most of the current research in this field is devoted to product reviews from websites. Microblogs and social networks pose even a greater challenge to sentiment classification. However, especially marketing and political campaigns leverage from opinions expressed on Twitter or other social communication platforms. The objects of interest in this paper are the presidential candidates of the Republican Party in the USA and their campaign topics. In this paper we introduce the combination of the noun phrases' frequency and their PMI measure as constraint on aspect extraction. This compensates for sparse phrases receiving a higher score than those composed of high-frequency words. Evaluation shows that the meronymy relationship between politicians and their topics holds and improves accuracy of aspect extraction.

## Introduction

Searching for people's opinions via surveys and polls has been an expensive and time-consuming task.

The proliferation of Web 2.0 has changed the way people express their opinions and feelings. This so called user-generated content posted in blogs, forums, product review sites and social networks is mostly publicly available and easy to obtain. The high value of this content arises from its subjective nature which, in aggregated form, indicates public opinion. It is difficult for humans to read and summarize all relevant documents in terms of the expressed sentiment. Thus, there is a growing need for automated analysis of this kind of data. This is a challenging task with foundations in natural language processing and text mining referred to as sentiment analysis. Many research studies in sentiment analysis are concerned with product reviews from websites like

Amazon or Epinions. Most of them propose supervised machine learning techniques trained on bag-of-words features to classify sentiment. Therefore, these approaches tend to be very domain specific and require large annotated corpora. Systems dealing with Microblogging services like Twitter and other social communication platforms are still in the early stage of development, although many applications, including politics and marketing campaigns, can leverage from the use of social media and especially Twitter. For example, online organizing techniques played an important role in the U.S. President Barack Obama's campaign in 2008. Current TV ran a program during the debate between John McCain and Barack Obama called "Hack the Debate" where they asked the audience to post comments on Twitter. After the success of Barack Obama, Twitter has become a legitimate communication channel in the political area. Therefore, it can be expected that sentiment analysis will be part of every campaign in the future. However, currently used systems in the industry are not capable of detailed opinion summaries such as distributions of opinions by topic or entity.

This paper studies the application of the Pointwise Mutual Information (PMI) measure to extract relevant topics of the campaign from tweets mentioning the names of the candidates of the Republican Party in the USA and their associated sentiment. The main contributions of the paper are:

- We introduce the combination of the noun phrases' frequency and their PMI measure as constraint on aspect extraction. (Constraint means that the value of this measure determines relevant aspects.) That way the problem that bigrams composed of low-frequency words will receive a higher score than those composed of high-frequency words can be compensated
- We show that the meronymy relationship between politicians and their campaign topics holds and adds valuable information to aspect extraction

## Roadmap

The paper is organized as follows.

In Section 2, collecting, storing and preprocessing the Twitter data using techniques which are part of the Natural Language Toolkit (NLTK) are described. Section 3 shows the aspect-extraction algorithm and the obtained results. Section 4 discusses first the used lexicon and its expansion as well as the sentiment classification on both word and aspect-level. The summary of opinions is given in Section 5. Section 6 contains evaluation of adjusted PMI measures and the results. Finally, we draw conclusions and examine possibilities of future work.

## Related Work

There are several research papers discussing sentiment analysis.

In his paper, Turney applied simple unsupervised learning algorithm for classifying reviews as recommended or not (Turney 2002). The classification of a review is predicted by the average semantic orientation of the phrases in the review that contain adjectives or adverbs. In contrast to it, in (Pang, Lee and Vaithyanathan 2002) is shown that the three standard machine learning techniques do not perform as well on sentiment classification as on traditional topic-based categorization. They also obtained better performance with word presence compared to frequency. This shows that, in contrast to standard text categorization, frequent occurrences of keywords are less important than their mere presence. The paper (Liu 2010) gives a general overview of the problem of sentiment analysis as well as of classification techniques and opinion search.

There are many papers, which describe different classification techniques for sentiment analysis. Sentiment classification can be formulated as a supervised problem with two class labels (positive and negative). In (Pang, Lee and Vaithyanathan 2002), the authors apply supervised learning methods such as naïve Bayesian and support vector machines (SVM) to classify movie reviews into two classes.

Most unsupervised sentiment classification approaches try to generate a general or domain dependent opinion lexicon for words or opinion phrases. In (Riloff and Wiebe 2003), the authors collected subjectivity clues as a part of their work. The clues were then used in (Wiebe, Wilson and Cardie 2005) to detect semantic orientation. In this paper, a bootstrapping process was developed, where high-precision classifiers use known subjective vocabulary to separate subjective and objective sentences from a non-annotated text collection. The aspect extraction method refers to the concept of determining opinion targets and their attributes which are mentioned in a document or a sentence. Many information extraction techniques have

been applied so far. In (Hu and Liu 2004), association mining was used to find frequent nouns and noun phrases as itemsets. Frequent itemsets mentioned in product reviews are believed to be aspects of the products. A more detailed strategy is to model a word order using class sequential rules (Hu and Liu 2006). Based on the assumption that aspects and opinion words often occur together in particular syntactic patterns, a method called double propagation aims to simultaneously extract new opinion words and aspects (Qiu, Liu and Chen 2011). A small pre-defined opinion lexicon and dependency trees are used to extract adjectives and directly related nouns and propagate this information back and forth. This method outperforms both state-of-the-art lexicon expansion and aspect extraction algorithms.

Sentiment analysis on Twitter data is considerably different from standard sentiment analysis techniques. General opinion lexicons are not able to cover the informal language of social communication platforms. Special regards to preprocessing are inevitable due to some unique properties of the Twitter language. The nature of Twitter messages, called tweets, cannot be compared to the structure of the current studies' domains. In (Pak and Paroubek 2010), tweets containing emoticons are used as training corpus to avoid manual annotation. They split up the data so that happy emoticons form the positive annotated set and sad emoticons form the negative annotated set. Their results show that along SVM and CRF (conditional random fields), Naïve Bayes classifier performed best and bigrams outperformed unigrams as features. One weakness of the automatic labeling process is that it is vulnerable to irony and sarcasm, which are often implied in emoticons. A similar method was proposed earlier in (Go, Bhayani and Huang 2009). They found that POS features decrease performance and bigrams are not useful due to data sparseness. Their baseline uses the public available list of positive and negative sentiment words from Twitter and assigns each tweet the polarity of the higher count of sentiment words of each class. One of the difficulties about analyzing text contained in tweets is the 140 character constraint on the message length. This forces Twitter users to communicate through a lot of shortcuts, slang and misspellings. In (Brody and Diakopoulos 2011), the authors present an automatic method which leverages word lengthening to adapt a sentiment lexicon specifically for Twitter and other social messaging networks. A different approach is made in (Davidov, Tsur and Rappoport 2010), where several Twitter tags and smileys have been used as sentiment labels to build a classification framework. The authors use different feature types (punctuation, words, n-grams and patterns) for classification and show that the framework successfully identifies sentiment types of the untagged tweets.

Sentiment analysis in politics is another important issue. Online informal political discussions in blogs have become an important feature of the intellectual landscape of the internet. In (Mullen and Malouf 2006), the authors collected posts from political blogs in order to classify the political orientation of authors as either left or right. They used the Naïve Bayes text classifier and conclude that traditional word-based text classification is inadequate for political sentiment analysis. (Diakopoulos and Shamma 2010) illustrates the usage of Twitter messages posted in conjunction with the last live presidential debate in the USA. They aggregate the overall sentiment on a timeline to see how particular topics affected the audience. Other studies, such as (Kim and Hovy, 2004), try to classify predictive opinions in election forums into „likely to win“ and „unlikely to win“. This task is closely related to sentiment classification, since the two opposing categories are subjective judgments.

### Collecting, Storing and Preprocessing Data

The objects of interest in this paper are the presidential candidates of the Republican Party in the USA and their campaign topics.

Tweets were queried several times in November and December of 2011 to obtain information mentioning the candidate names. Only active candidates who were invited to the primary debate at the Oakland University in Rochester, Michigan on November 9, 2011 and the following debates were chosen as relevant. Table 1 shows the list of candidates and the number of tweets in relation to each of them.

Candidate	Number of tweets
Herman Cain	7204
Jon Huntsman	5741
Mitt Romney	7243
Michele Bachmann	6864
Newt Gingrich	7199
Rick Santorum	5932
Ron Paul	7279
Rick Perry	7266

Table 1: Republican presidential candidates and the number of corresponding tweets

Twitter offers two APIs to retrieve data from tweets: REST and Streaming. For this paper, the REST API was used and called via Python. Tweets were retrieved in form of the

Javascript Object Notation (JSON) documents. NoSQL document database systems offer convenient functionality to store JSON objects. Therefore, the tweets for the aim of this paper were stored in the mongoDB database system. A database in mongoDB is organized in collections, while a collection holds documents. Tweets of each political candidate were stored in different collections. Extracted aspects and their sentiment were later inserted into a relational MySQL database, to allow a presentation of results to users.

For the purposes of natural language processing Python offers the Natural Language Toolkit (NLTK) module. NLTK provides functionality for symbolic and statistical natural language processing, including data preprocessing. In the first step, the NLTK built-in Punkt Sentence Tokenizer was used to divide a text into a list of sentences, without destroying the integrity of words and sentences. After that, the Treebank Word Tokenizer splits an input sentence into single tokens. Finally, the NLTK’s tagger is used for POS tagging of items. In relation to noun phrases, a regular expression parser is applied to determine noun phrase chunks in a given pos-tagged sentence.

The following tokens of the tweets were removed in preprocessing:

- Hyperlinks and hashes (tags were preserved)
- Usernames with preceding “@” which expresses a reply to said user
- The “RT” keyword which is used to indicate that the following tweet is a retweet

### Aspect Extraction

#### Methods for Extraction

In this section, the methods which were used to extract aspects out of the gathered tweets are described.

Extracted noun phrases are assumed to be campaign topics and are measured by the correlation to the corresponding candidate via PMI. The PMI measure and phrase frequency form a constraint on the subsequent analysis of their sentiment. According to (Liu 2010), during aspect extraction (also called feature extraction) it is assumed that only noun phrases are relevant aspects of the opinion targets. In this work the targets are political candidates and important topics in their campaigns are believed to be noun phrases. The usage of noun phrases has one shortcoming, since it extracts only explicit aspects which the opinion holder expresses directly. Other phrases like verbs may indicate implicit aspects. Implicit aspects will not be considered in this work. Phrase structures are usually represented as a tree, where the leaf nodes are words and intermediate nodes are phrases. The tree has a single root node, which is the start symbol of the grammar.

Not every extracted noun phrase can be seen as a relevant aspect. For this reason we need to prune the set of all phrases. A common assumption is that, given enough data, the vocabulary of the corpus will converge. Hence, noun phrases which occur more frequently are more likely to be an aspect. In (Hu and Liu 2004), the authors define an itemset as frequent if it appears in more than 1 % of the sentences. This approach has one shortcoming, as infrequent noun phrases may contain important aspects, but they are instantly pruned. In this work aspects are pruned on a combination of frequency and PMI measures in order to cover both the importance of frequent and infrequent aspects having a strong correlation to the object. In (Popescu and Etzioni 2005), the authors proposed PMI statistics to improve aspect extraction on product reviews. The paper states that a product name and its aspects form a meronymy. (A meronymy is a part-whole relationship, e.g. the word “tire” is a meronym of “car”.) The authors calculate the PMI measure between the candidate aspect phrase and some meronymy discriminators associated with the product class. In our work we assume that important topics of political election candidates also form a meronymy, because the topics they talk about are parts of their campaigns representing the candidate. The meronymy discriminators for the work in this paper were chosen experimentally, and the set  $\{of, 's, about\}$  provided the best results. The authors of (Manning and Schütze 1999) argue that sparseness is a difficult problem for PMI, because bigrams composed of low-frequency words will receive a higher score than those composed of high-frequency words. (The notion of Pointwise Mutual Information has been introduced in (Dunning 1993).) This problem is compensated in our work, as can be seen in the next subsection. Equation 1 shows the simplified version of the PMI measure used to search the web for the number of hits of phrases  $x$  and  $y$ , which is often called PMI-IR (Information Retrieval).

$$PMI_{IR}(x, y) = \frac{hits(x \text{ NEAR } y)}{hits(x)hits(y)} \quad (\text{Eq. 1})$$

The numerator in Equation 1 determines the number of hits which contain both phrase  $x$  and phrase  $y$ , where  $x$  stands for the candidate name plus a concatenated discriminator and  $y$  is the candidate aspect. To make this description more specific, the NEAR-Operator of the Bing search engine was utilized. This operator makes sure that the distance between phrase  $x$  and phrase  $y$  is less than a maximum distance. Therefore, the returned number of hits is more accurate. In this work the maximum distance was set to ten words. For instance, the PMI measure for candidate Mitt Romney and the aspect “teaparty” is calculated for each discriminator from the set  $\{of, 's,$

$about\}$  and is given in Equation 2. ( $d$  denotes the current discriminator)

$$PMI_{IR}(Mitt Romney + d, teaparty) = \frac{hits(Mitt Romney + d \text{ NEAR } teaparty)}{hits(Mitt Romney + d)hits(teaparty)} \quad (\text{Eq. 2})$$

### Constraint on Aspect Extraction

As we already stated, sparseness is a difficult problem for PMI, because bigrams composed of low-frequency words will receive a higher score than those composed of high-frequency words.

The problem of sparseness is compensated so that a combination of the noun phrases’ frequency and their PMI measure operates as constraint on aspect extraction. The top resulting aspects of each candidate sorted with respect to the product of the average PMI measure ( $\mathbf{PMI}_{avg}$ ) and the adjusted count of tweets ( $\mathbf{c}_{adjust}$ ) represent important topics. Different weighting combinations of the two factors were tried, but did not improve the results. In Table 2, the  $\mathbf{c}$  denotes the overall count of tweets mentioning the aspect, while  $\mathbf{c}_{rt}$  is the fraction of retweets. In this work, a retweet is of lower value than an original tweet, because it is less effort for Twitter users to just retweet something they approve than to think of their own comments. It is assumed that the spread of retweets has exponential growth. (In other words, we assume that retweets and their “offspring” form a tree.) Equation 3 shows the calculation of the adjusted count of tweets.

$$c_{adjust} = (c - c_{rt}) + \ln(c_{rt}) \quad (\text{Eq. 3})$$

Table 2 shows that the mistakes in POS tagging were made regarding the phrases “company showed profits”, “the character of his opponents” and “defends”. Furthermore, with phrases “cnn”, “washington post”, “boston globe” and “reuters” there are four news media achieving high PMI measures. This indicates that many online articles about the candidate have been published. In general, it is difficult to tell if the extracted aspects are political topics, or at least a part of them without any provided context. For example:

IT’S ON: Democrats Are Waging War Over Romney’s ‘Sleazy’ Campaign Ad [http://...](#)

More accurate results are expected with the use of more reliable POS tagging and by gathering a bigger corpus to improve vocabulary convergence. A bigger corpus is



<b>c</b>	<b>c<sub>rt</sub></b>	<b>Aspect</b>	<b>c<sub>adjust</sub></b>	<b>PMI<sub>avg</sub> × c<sub>adjust</sub></b>	<b>c<sub>pos</sub></b>	<b>c<sub>neg</sub></b>	<b>c<sub>neu</sub></b>
36	19	planned parenthood	19.94	4.69	6.00	13.94	0
218	96	teaparty	126.56	4.01	48.76	47.14	36.40
36	20	boston globe	19.00	3.76	15.71	4.61	0
16	3	the character of his opponents	14.10	3.19	0	14.10	0
77	7	campaign ad	71.95	3.17	1.00	69.95	1.00
11	4	an equality advocate out	8.39	2.86	0	8.39	0
153	80	mitt2012	77.38	2.40	29.40	27.50	25.83
67	6	defends	62.79	2.14	2.00	60.79	0
9	2	job prospects	7.69	2.04	7.69	0	0
55	5	obama quote	51.61	2.00	0	51.61	0
9	8	a repeat	3.08	1.98	0	3.08	0
52	22	sesame street	33.09	1.83	18.39	15.89	0
96	43	a beer	56.76	1.82	0	56.76	0
25	0	reuters	25.00	1.82	11.00	14.00	0
68	23	a cigarette	48.14	1.76	0	48.14	0
56	6	president Obama	51.79	1.69	11.10	41.10	0
59	22	wants	40.09	1.68	9.71	30.95	1.00
9	1	cnn poll	8.00	1.67	4.00	0.00	4.00
76	30	company showed profits	49.40	1.65	0	49.40	0
10	0	washington post	10.00	1.55	0	9.00	1.00

Table 2: Mitt Romney: Aspect extraction results

crucial here, because a closer look on different tweets responsible for these results showed that there are a lot of identical tweets which are not marked as retweets. Most of them are messages containing the headline of online news articles posted by different users. This leads to biased counts and unreliable results.

## Sentiment Classification

Usually, unsupervised sentiment classification has two steps.

First, a general or domain dependent opinion lexicon for words or opinion phrases is generated. After that, sentiment is classified based on a statistical measure.

### Building Lexicon

Most unsupervised sentiment classification approaches try to generate a general or domain dependent opinion lexicon for words or opinion phrases.

In our work, the subjectivity clues lexicon, which was presented in (Wilson, Wiebe and Hoffmann 2005), was used to detect semantic orientation. The resulting subjective clues were later annotated with their prior polarity using different manually developed sources and consist of 2296 positive, 4138 negative and 444 neutral distinct opinion words available at:

[http://www.cs.pitt.edu/mpqa/subj\\_lexicon.html](http://www.cs.pitt.edu/mpqa/subj_lexicon.html)

Most established sentiment lexicons, as one discussed above, were created considering no particular domain and

suffer from limited coverage and inaccuracies when applied to the highly informal domains like tweets or social networks communication. In (Ringsquandl and Petkovic 2012), we extract domain specific adjectives from the Twitter corpus and expand the general lexicon based on the work in (Hatzivassiloglou and McKeown 1997). Also, in this paper we examine a semi-supervised approach to expand the general lexicon with domain specific opinion words. This approach is based on the idea of sentiment consistency and can easily be transferred to other domains.

### Lexicon-Based Classification

In this section we discuss word-level and aspect-level sentiment.

#### Word-Level Sentiment

We assume that semantic orientation of word  $w$  is the class which maximizes the probability  $c$  conditional on  $w$ , where  $C = \{positive, negative, neutral\}$  and  $c \in C$ .

Every word  $w$  can be represented as the set of its synonym retrieved from WordNet. Equation 4 presents the final derived formula used for polarity classification. The denominator of Bayes' theorem is ignored as it is a normalizing constant.

$$\begin{aligned}
 SO(w) &= \arg \max_{c \in C} P(c|w) \\
 &= \arg \max_{c \in C} P(w|c) P(c)
 \end{aligned}$$

$$\begin{aligned}
&= \arg \max_{c \in C} P(\text{syn}_1, \text{syn}_2, \dots, \text{syn}_n | c) P(c) \\
&= \arg \max_{c \in C} \frac{\sum_{i=1}^n \text{count}(\text{syn}_i, c)}{|\text{synset}_w|} P(c)
\end{aligned} \tag{Eq. 4}$$

For every synonym,  $\text{count}(\text{syn}, c)$  is 1, if the synonym appears with polarity  $c$  in the opinion lexicon, otherwise it is 0. If no synonyms can be found, semantic orientation of the word  $w$ , denoted by  $SO(w)$ , is the prior polarity of word  $w$  in the opinion lexicon. Polarity of a word is adopted from the opinion lexicon without regards to its POS tag, i.e. it is assumed that prior polarity has same orientation independent of its POS tag. (Words that cannot be found in the augmented opinion lexicon are assumed to have neutral polarity.) For instance, the word “small” has eight synonyms, three of them are positive, one is neutral and four are negative, according to the opinion lexicon.

$$\begin{aligned}
P(\text{positive}) &= 0.34, P(\text{negative}) = 0.60, \\
P(\text{neutral}) &= 0.06
\end{aligned}$$

$$\begin{aligned}
SO(\text{small}) &= \arg \max_{c \in C} \left\{ \frac{4}{8} 0.6, \frac{3}{8} 0.34, \frac{1}{8} 0.06 \right\} \\
&\Rightarrow SO(\text{small}) = \text{negative}
\end{aligned}$$

Directly preceding negations change the semantic orientation from negative to positive and vice versa. Neutral sentiment is not affected by negations.

#### Aspect-Level Sentiment

The final aspect-level sentiment is determined by a simple aggregation function which sums up the semantic orientation of all words in the sentence  $s$  that mentions aspect  $a$ .

Every semantic orientation is weighted relative to its distance to the aspect. The distance of the current word and the aspect is the number of words lying in between. The idea behind this function is that opinion words which are close to the aspect are most likely to be related to it. Equation 5 shows how the aspect-level sentiment score has been calculated in this paper. ( $\text{score}(a, s) > 0$  means that a sentiment about the aspect is positive,  $\text{score}(a, s) < 0$  means negative,  $\text{score}(a, s) = 0$  is neutral.)

$$\text{score}(a, s) = \sum_{w_i \in s} \frac{SO(w_i)}{\text{dist}(w_i, a)} \tag{Eq. 5}$$

### Opinion Summaries

Aggregated aspect-level sentiment for each candidate is visualized in form of a bar chart.

The horizontal axis denotes the top 20 phrases and the vertical axis represents the adjusted number of tweets subdivided into polarity classes. The number of tweets is adjusted accounting for a logarithmic weight of retweets as discussed in Section 3.

Figure 1 presents the final opinion summary for Mitt Romney. The chart shows that mostly negative sentiment is expressed on the extracted topics. It also points out that the aspects “teaparty” and “mitt2012” which are often used as hashtags, exhibit uniform class distribution of all polarity classes. (Hashtags are short keywords used to indicate the topic of the tweet or the attitude of the user.) In this work, only the hash character is removed and the tag itself is preserved for later analysis. Therefore, hashtags occur in various tweets and might not be useful for the extraction of important topics. What can be seen without any further analysis is that apparently, Mitt Romney’s campaign ad and his quote of president Obama did not achieve a good resonance of the Twitter users.

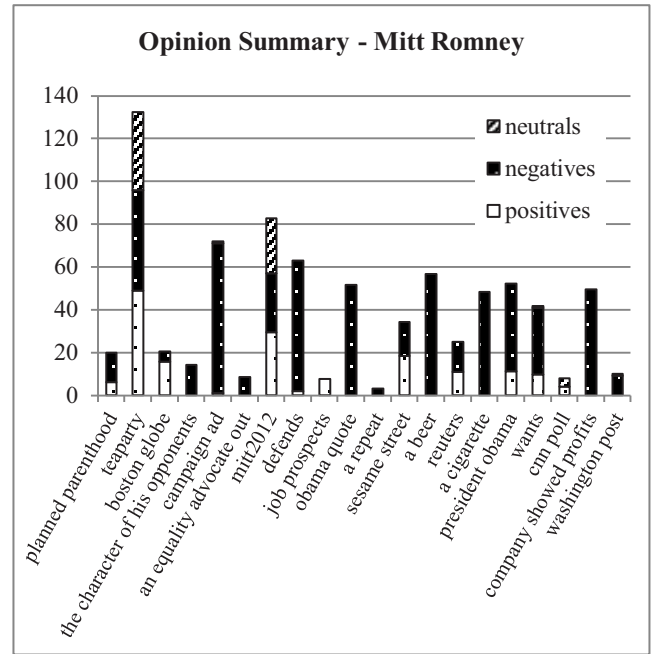


Figure 1: Opinion summary for Mitt Romney

In the summary of Rick Santorum shown in Figure 2, the topics about “gay marriage” and “muslims”, for example, were discussed controversially with an overweight of negative opinions. However, most topics are difficult to interpret without any provided contextual information. Hence, information systems using such an opinion summary should allow users to drill-through the aggregated data and present a list of distinct messages for each topic of interest, i.e. a mixture of text-based summaries and quantitative visualization techniques.

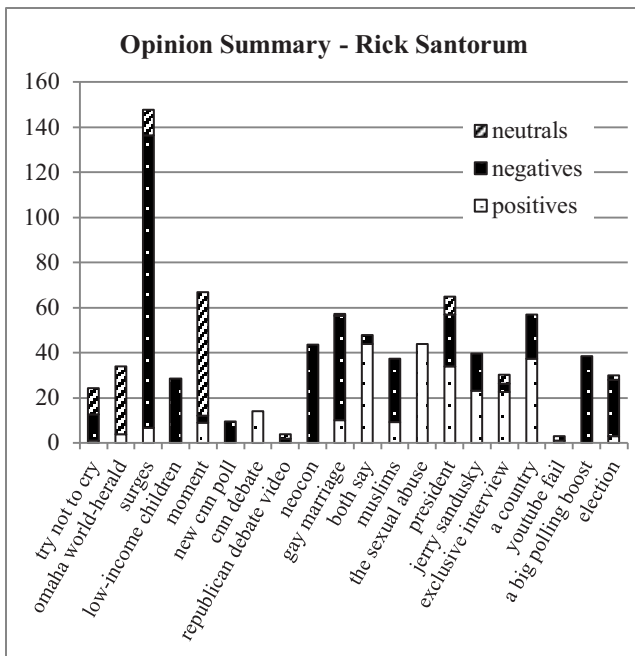


Figure 2: Opinion summary for Rick Santorum

## Evaluation and Results

In order to evaluate the performance of the adjusted PMI measure (the product of the average PMI measure and the adjusted count) as constraint on aspect extraction we annotated all extracted noun phrases of two political

candidates either as “aspect” or “no aspect”.

A noun phrase is labeled as aspect if it represents either a generic political topic, e.g. “foreign policy”, or a concrete topic that was relevant for this election’s context, e.g. “occupy movement”. Classification of noun phrases is based on their constraint score. A higher score means that the noun phrase is more likely to be an aspect. Figure 3 compares the performance of the adjusted PMI measure to pure frequency score as constraint on aspect extraction.

The lift chart in Figure 3 visualizes classification performance depending on the number of included noun phrases in Figures 1 and 2. Note that in the critical region located between 0 and 3 percent noun phrase ratio, where the highest scoring noun phrases are located, adjusted PMI measure correctly classifies more aspects than frequency-based scoring. Later, we set the threshold of included noun phrases to 20. Table 3 justifies this setting, as it presents average classification accuracies that were calculated on two different data sets (Rick Perry and Mitt Romney) with varying threshold of included noun phrases. In our case, the adjusted PMI measure as constraint achieves highest average accuracy at a threshold of 20. These results can be interpreted as follows: The PMI adjustment weights out the score of some of the frequent phrases that are, although high frequency-based score, no aspects, and tries to give low-frequency aspects a scoring boost. This leads to more accurate extraction results than pure frequency scoring. Additionally, this implies that the meronymy relationship between politicians and their campaign topics holds.

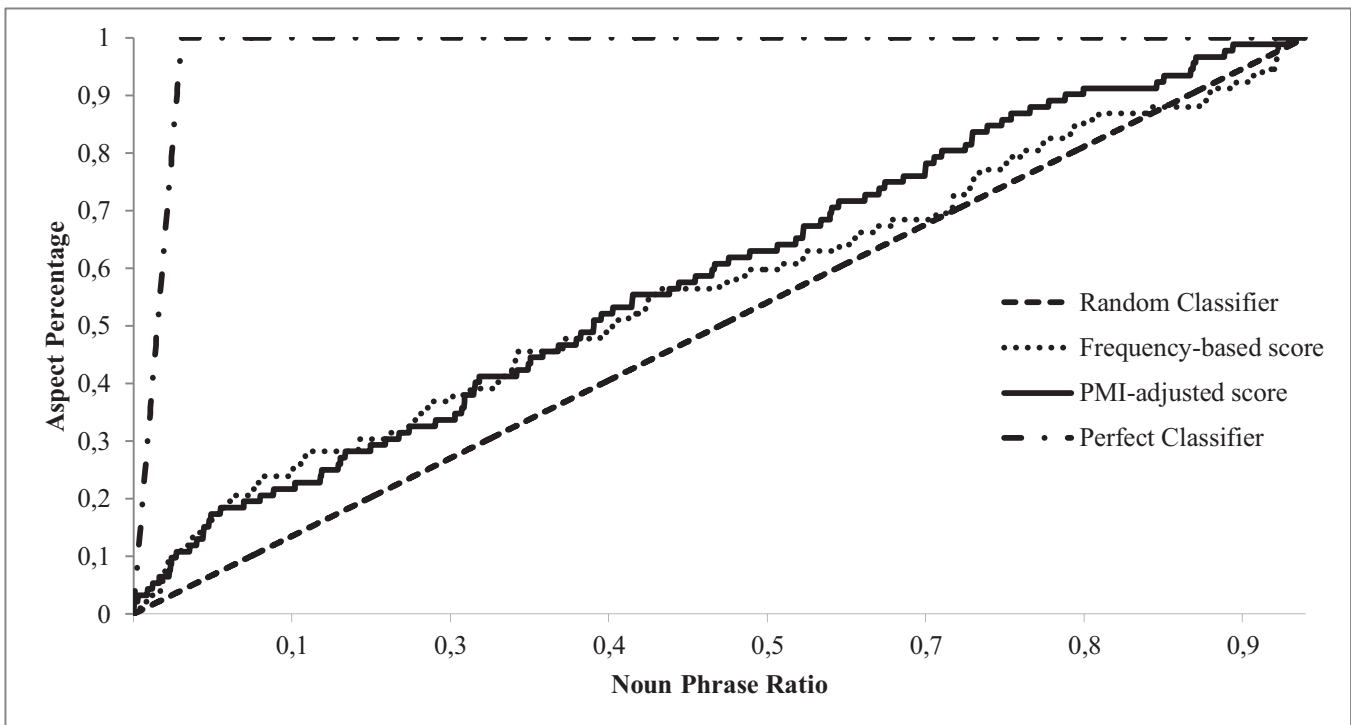


Figure 3: Lift chart for classification performance

Threshold	Average accuracy	
	Adjusted PMI measure	Frequency score
20	<b>22,5%</b>	17,5%
40	20,25%	16,25%
60	17,50%	18,34%
80	16,75%	<b>20,00%</b>
100	16,00%	19,50%

Table 3: Average accuracies

## Conclusions and Future Work

This paper presents the challenging task of aspect-based opinion summarization on Twitter data in the domain of politics, which falls into the application category of social media monitoring.

It was discussed that although Twitter data can easily be gathered, special considerations in retrieval and pre-processing are needed. NLTK's built-in pre-processing functionalities were found to be not completely sufficient for informal text corpora. We extracted relevant aspects with a newly introduced combination of the PMI measure and phrase frequency as constraint. Aspect extraction and pruning methods presented in this paper can be applied in any domain where a meronymy relationship of opinion targets and aspects holds true. We verified this relationship for political candidates and their campaign topics. The evaluation of the PMI adjusted measure as constraint on aspect extraction shows that the meronymy relationship between politicians and their campaign holds. Possibilities for future work include the learning of other domain-specific opinion words like nouns and verbs. Such a classification task would probably need to involve syntactic dependencies as features. Both time and regional distinctions could reveal trends and allow a more detailed presentation of political topics and associated sentiment. This could reveal that a certain topic causes positive reactions in one state, while it gets mostly negative comments in another state. In terms of aspect-level sentiment, the simple distance-weighted score presented here can be improved when it is assured that particular opinion words are expressed in relation to the aspect or the opinion target. Sophisticated analysis of long-distance opinion shifter dependencies are expected to increase the reliability of aggregated aspect sentiment.

## References

Brody, S. & Diakopoulos, N., 2011. Using Word Lengthening to Detect Sentiment in Microblogs. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 562-570.

Davidov, D., Tsur, O. & Rappoport, A., 2010. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys.

*Proceeding of the 23rd international conference on Computational Linguistics (COLING)*.

Diakopoulos, N. & Shamma, D., 2010. Characterizing Debate Performance via Aggregated Twitter Sentiment. *Proceedings of the 28th int. conference on Human factors in computing systems*.

Go, A., Bhayani, R. & Huang, L., 2009. Twitter Sentiment Classification using Distant Supervision.

Hatzivassiloglou, V. & McKeown, K., 1997. Predicting the Semantic Orientation of Adjectives. *Proceedings of the Joint ACL/EACL Conference*, pp. 174-181.

Hu, M. & Liu, B., 2004. Mining and Summarizing Customer Reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Hu, M. & Liu, B., 2006. Opinion Feature Extraction Using Class Sequential Rules. *Proceedings of the Spring Symposia on Computational Approaches to Analyzing Weblogs*.

Kim, S.-M. & Hovy, E., 2004. Determining the Sentiment of Opinions. *Proceedings of the International Conference on Computational Linguistics*.

Liu, B., 2010. Sentiment Analysis and Subjectivity. In: *Handbook of Natural Language Processing*. Boca Raton, FL: CRC Press, pp. 627 - 666.

Manning, C. & Schütze, H., 1999. *Foundations of Statistical Natural Language Processing*. Massachusetts Institue of Technology - Cambridge: THE MIT PRESS.

Pak, A. & Paroubek, P., 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh conference on International Language Resources and Evaluation*.

Pang, B. & Lee, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*.

Pang, B., Lee, L. & Vaithyanathan, S., 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Popescu, A.-M. & Etzioni, O., 2005. Extracting Product Features and Opinions from Reviews. *Proceedings of Conference on Empirical Methods in Natural Language*.

Qiu, G., Liu, B., Bu, J. & Chen, C., 2011. Opinion Word Expansion and Target Extraction using Double Propagation. *Computational Linguistics Vol. 37*.

Riloff, E. & Wiebe, J., 2003. Learning Extraction Patterns for Subjective Expressions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Ringsquandl, M. & Petkovic, D., 2012. Expanding General Lexicon with Domain Specific Opinion Words Using Semi-Supervised Approach. *BRACIS 2012 - WTI, IV International Workshop on Web and Text Intelligence*.

Turney, P., 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the Association for Computational Linguistics*, p. 417-424.

Wiebe, J., Wilson, T. & Cardie, C., 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, pp. 164-210.

Wilson, T., Wiebe, J. & Hoffmann, P., 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*.