

# Polarity Detection of Sarcastic Political Tweets

D. K. Tayal, Sumit Yadav, Komal Gupta, Bhawna Rajput, Kiran Kumari

Department of Computer Science and Engineering,  
Indira Gandhi Delhi Technical University for Women,  
Kashmere Gate, Delhi, India.

E-mail: [devendratayal@igit.ac.in](mailto:devendratayal@igit.ac.in)

**Abstract** - Sarcasm is an activity of saying or writing in such a way that the textual meaning of what is said is opposite of what is meant. Generally, sarcastic sentences are used to express negative feelings. Thus in political polarity detection, it can be used to know the dissatisfaction of people against a particular party, candidate or government and hence considering this other aspects can determine the poll results. This paper presents two algorithms, one to identify a sarcastic tweet and other to perform polarity detection on political sarcastic tweets. It also includes the current approach we are using to identify the sarcastic political tweets and future aspects. It also emphasizes on using the social networking sites and mainly Twitter as a source for predicting political results.

## I. INTRODUCTION

Indirect speech or text that conveys an implicit meaning is called SARCASM. Because of this characteristic of sarcasm it is hard to recognize sarcasm. The difficulty in recognition of sarcasm causes misunderstanding in everyday communication. In addition to identification, polarity detection of sarcasm is also a complex task and adds to the complexity in determining polarity. This is the reason not much work is done on polarity detection through sarcasm.

According to Macmillan English Dictionary (2007) Sarcasm - "the activity of saying or writing the opposite of what you mean, or of speaking in a way intended to make someone else feel stupid or show them that you are angry".

The goal of our research is to analyze and predict which party and/or candidate will win in the upcoming Indian Central Government Election of 2014 based on sarcastic tweets. The most important goal is to find out whether more number of sarcastic tweets for a particular party or candidate can result in lesser number of seats in central government elections. We have chosen Twitter as a media to compute the poll central government elections results because Twitter has a constraint of 140 characters on the length of the tweet. This gives shorter tweets which are easy to analyze and provide crisp and clear opinion of the author. Also twitter hash tag assists in collecting datasets as it uniquely categorizes twitter corpus.

## II. RELATED WORK

A good amount of work has been done in identification of sarcasm. Some of the examples are: Davidov [7] proposed a supervised classification framework for analyzing sentiments in twitter data. They used the hash tags given along the tweets and four feature types for sentiment classification: single word features, n-gram features, pattern features and punctuation features. Davidov [8] used semi-supervised method to identify sarcastic sentences. They created a gold sample which consists of tags corresponding to each sentence tagged by 3 annotators. Karamibekr [1] described the importance of role of verb in sentiment analysis of social issues. Experimental results shows that verb is not only required but it also improves the performance of sentiment analysis. Nooralahzadeh [13] compared the sentiments before and after the presidential elections, held in both US and France in the year 2012 and extracted the tweets from Twitter, scored the sentiments in each tweet using different metrics and performed a time series analysis for candidates.

## III. APPROACH

The entire process includes following steps: 1) Collecting the twitter dataset. 2) Data Pre-processing: It includes removing URLs and references to other users which are of the form @user.

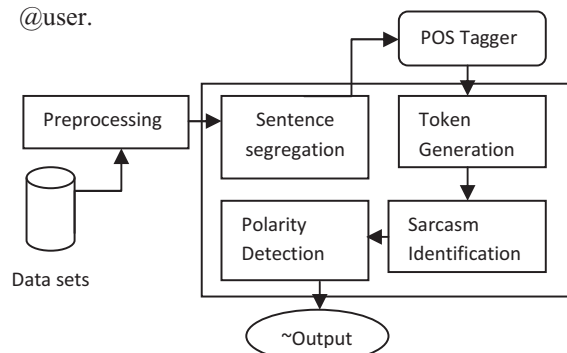


Figure 1

3) Sarcasm Identification 4) Determining the polarity at phrase level, sentence level 5) Polarity Detection of the result. 6) Prediction of election results. The above figure is the diagrammatic

representation of our approach. The final output which we get above is the polarity of the sarcastic sentence. Note the tilde (~) sign in the output of the above diagram which shows that the output should be reverse analysed because in sarcasm anything which is to be said is commented in a way exactly opposite to what the fact is. So the polarity of the above output should be reversed.

The equation which we are using to find the polarity of a sentence is:

$$\text{polar}_s = \text{polar}_s + \text{polar}_j \quad (1)$$

Where  $\text{polar}_j$  is the polarity of an opinion adjective and verb along with consideration of negations included like don't, not etc. which reverses the impact of opinion these adjectives and verbs can impose.

The polarity for a tweet is then calculated by using the below equation:

$$\frac{\sum_{i=1}^n (S_i)}{\text{Number of sentence in the document}} \quad (2)$$

This will calculate the overall polarity of tweets regarding a particular candidate.

#### A. Verb and Adjective analysis

Since political sarcastic tweets mainly shows either comparison between two candidates which is a way of direct disagreement to one of them or indirect disagreement to a comment or behaviour shown by them. For a general sentence to be sarcastic; adjectives with the highest positivity along with some punctuation marks like exclamation sign or more number of question marks directs a high probability for the sentence to be sarcastic. For example "Minister's poverty and their so called intelligence exist in parallel universe only... ;)". Here the verb "exist" and the adjectives "poverty" and "intelligence" shows that imposing these superior/inferior adjectives to the verb gives a different meaning to its direct object (candidates). Moreover the use of the emoticon like ";" used at the end of this example strengthens the fact that it is a sarcastic statement. To recognize the elements of a sentence we have employed Stanford POS tagger [1]. This tagger tokenizes the sentence and classifies each word as part of speech that is as an adjective, noun, pronoun, verb, determinant etc. With this we can well identify the noun, verbs and adjective and can determine the relation of the adjective and verb with the direct subject or noun used in the tweet.

#### B. Opinion Dictionary

Our approach is based on the lexical analysis by extracting the opinion adjective and verbs in the sentence. Also since it is the analysis of sarcastic sentence so we have also employed a dictionary where all the included words have been assigned a polar strength using which we have calculated overall polarity of each sentence. So using it we

can calculate the overall polarity of the sentence as well as document.

#### C. Training Set

Since we are dealing with sarcasm where the actual meaning of the sentence is exactly opposite to the lexical meaning of the sentence, so for identifying sarcasm in political tweets one must have prior universal information about the candidates and parties. For example, if a politician is having allegations of corruptions, then we can easily conclude that a tweet mentioning that the politician is an honest person is making sarcasm onto him. So, our training set contains some information or adjectives related to one of the candidate/politician/political party and their various antonyms so that we can compare the adjectives used in each tweets with these antonyms to improvise our efficiency of detecting sarcastic tweets. This idea evolves from the fact that people suffering from memories pertaining problem finds it difficult to identify sarcasm in a sentence. So for a person to identify a sarcastic sentence, it is much necessary for him/her to have prior persistent knowledge about the fact for which the comment is made. For example "Congress is for Good Governance! #Sarcasm". In this sentence what if the hash tag and "!" (Exclamation) were not there? So for better understanding of sarcasm, knowledge of the actual fact should be made available. So the information list or training set which we are maintaining that basically contains words collected after analyzing a number of political sarcastic tweets.

#### D. Techniques

We are using supervised approach to identify the sarcastic sentence's polarity. We will use punctuation marks ! , ? characters in the sentence, #sarcasm tag and #irony, emoticons and adjective and verb in conjunction with ! in Twitter. Along with these, we also consider emoticons with verbs and adjectives since mere use of emoticons like ":P" and ";" can be used for humour also. Since most of the sarcastic tweets use hash tags but if these tags are missing then methods described in Section III is used to determine the sarcasm. We have designed an algorithm to take out adjectives and verbs along with considering their negation example usage of "not" etc; finding the polarity of sentence and finally calculating the polarity of the tweet and categorizing it into three of our assumed scales: negative, neutral and positive in the range -1 to 1. Range [-1,0) is classified as having negative polarity, 0 as neutral and (0,1] having positive polarity. We have also categorized the tweets as clearly sarcastic, probable sarcastic and non sarcastic. The polarity of the sarcastic tweets is calculated as follow-

---

*Algorithm 1: Polarity of sarcastic tweets.*

---

- 1: Split the document with tweets for a particular candidate into sentences.
  - 2: Extract the opinions from each sentence by recognizing the opinion adjectives, emoticons, opinion verb in conjunction with punctuation marks and hash tags.
  - 3: Determine the polarity of each opinion verb and adjectives regarding the political issue using polarity defined in the opinion dictionary.
  - 4: Calculate the overall polarity of each sentence using Eq.1.
  - 5: Calculate the polarity of each document using Eq. 2.
- 

The opinions used in a sentence are extracted (step 2) by implementing lexical analysis in each sentence using the Algorithm 2.

---

*Algorithm 2: To extract tokenized adjectives and verbs.*

---

1. Run the POS tagger to tokenize adjectives, verbs, nouns etc.
  2. Determine the adjectives, main verb and hash tags.
  3. Find the subject (Noun for which the verb is used) with the help of the main verb (step 2).
- 

*E. Example*

“NAMO is responsible for all da Riots, Corruption,HorseTrading,misuse of CBI, #PaidMedia, Terrorism etc. But Cong is for Good Governance! #Sarcasm”

*1)Sentence Segregation:*

Sentence1-“NAMO is responsible for all da Riots, Corruption,HorseTrading,misuse of CBI, #PaidMedia, Terrorism etc. “  
Sentence2-“But Cong is for Good Governance! #Sarcasm”

*2)Token Generation:*

a)*Sentence1:*“NAMO/NNPis/VBZ responsible/JJ for/IN all/DT da/NN Riots/NNS ./, Corruption/NNP ./, HorseTrading/NNP ./, misuse/NN of/IN CBI/NNP ./, #PaidMedia/NNP ./, Terrorism/NNP etc./FW ./.”

b)*Sentence2:* “But/CC Cong/NNP is/VBZ for/IN Good/JJ Governance/NN !/. #Sarcasm/NN”

Polarity of token good is 0.375 and good governance is compared with subject Cong, we come to know *Sentence2* is sarcasm hence its polarity is 0.375, and negating it gives -0.375. Similarly in *Sentence1* after summing the polarity of its tokens found in dictionary, we get polarity as -0.475. As we can observe the conjunction ‘but’ which is used to compare NAMO and Cong and we already identified *Sentence2* is used in a sarcastic way, therefore *Sentence1* is also a sarcasm on main

subject NAMO. Hence here also we will change the polarity of *Sentence1* and it will be +0.475. So we can conclude in this tweet *Sentence1* is having positive polarity that is for NAMO and *Sentence2* has negative polarity that is for Cong. If the document contains n sarcastic tweets about a candidate then the overall polarity will be calculated by using equation 2 that is by summing up all the polarities and dividing it by n.

#### IV. CONCLUSION AND FUTURE SCOPE

We have finally concluded that using the supervised approach and the above proposed algorithm we can achieve our goal i.e. identification of sarcastic tweets, determining polarity through it and predicting the results of some political results up to an efficient level. In future, efficiency and accuracy of an algorithm can further be increased by using Natural Language Processing more appropriately by considering the n-gram feature based dictionary.

#### REFERENCES

- [1]. Karamibekr, M.; Ghorbani, A.A., "Sentiment Analysis of Social Issues," in Proc. "ASE International Conference on Social Informatics" (SocialInformatics 2012), Washington, D.C, pp. 215-221, 2012.
- [2]. A. Bakliwal, J. Foster, J. Van Der Puil, R. O. Brien, L. Tounsi, and M. Hughes, "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier," in Proc. " Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies" (NAACL 2013), pp. 49–58, 2013.
- [3]. Bermingham, Adam, and Alan F Smeaton, "On Using Twitter to Monitor Political Sentiment and Predict Election Results," in Proc. "International Joint Conference for Natural Language Processing" (IJCNLP), pp. 2-10, 2011.
- [4]. Chen, Hailiang, Prabuddha De, Yu Hu, and Byoung-Hyoun Hwang, "Sentiment Revealed in Social Media and its Effect on the Stock Market," in Proc. "Statistical Signal Processing Workshop" (SSP), 2011 IEEE , pp.25-28, June 2011.
- [5]. J. Chung and E. Mustafaraj, "Can Collective Sentiment Expressed on Twitter Predict Political Elections?," in Proc. "25th AAAI Conf. on Artificial Intelligence" pp. 1770–1771, 2010.
- [6]. B. O. Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in Proc. "International AAAI Conference on Weblogs and Social Media"(ICWSM, 11), pp. 122-129, May, 2010.
- [7]. D. Davidov, Dmitry, O. Tsur, A. Rappoport, "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys," in Proc. "23rd International Conf. on Computational Linguistics:Posters", pp. 241–249, 2010.
- [8]. D. Davidov, Dmitry, O. Tsur, A. Rappoport, "Semi-supervised Recognition of Sarcastic sentences in Twitter and Amazon," in Proc. "14th Conf. on Computational Natural Language Learning", pp. 107-116, 2010.
- [9]. Elena, Filatova, "Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing," in Proc. "8th International Conf. on Language Resources and Evaluation" (LREC), pp. 392–398, 2012.
- [10]. González-ibáñez, Roberto, S. Muresan and N. Wacholder, "Identifying Sarcasm in Twitter: A Closer Look," in Proc. "49th Annual Meetings of the ACL:

- Human Language Technologies" (short papers, vol. 2), pp. 581–586, 2011.
- [11]. Hasan, S.M. Shamimul and D.A. Adjeroh "Proximity-Based Sentiment Analysis," in Proc. "4th International Conf. on Applications of Digital Information and Web Technologies" (ICADIWT), pp 106-111, August 2011.
  - [12]. D. Nasa, "Text Mining Techniques- A Survey," International Journal of Advanced Research in Computer Science and Software Engg., vol. 2, no. 4, pp. 50-54, 2012.
  - [13]. F. Nooralahzadeh, V. Arunachalam, and C.-G. Chiru, "2012 Presidential Elections on Twitter -- An Analysis of How the US and French Election were Reflected in Tweets," in Proc. "IEEE 19th Int. Conf. Control Syst. Comput. Sci.", pp. 240–246, May 2013.
  - [14]. A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in Proc. "7th International Conf. on Language Resources and Evaluation" (LREC'10), pp. 1320–1326, 2010.
  - [15]. V. K. Singh, R. Piryani, and A. Uddin, "Sentiment Analysis of Movie Reviews: A new feature-based heuristic for aspect-level sentiment classification," in Proc. "International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing" (IMAC4S), pp. 712–717, 2013.
  - [16]. V. K. Singh, D. Mahata, and R. Adhikari, "Mining the Blogosphere from a Socio-political Perspective," in Proc. "International Conf. on Computer Information Systems and Industrial Management Applications" (CISIM), pp. 365–370, 2010.
  - [17]. O. Tsur, D. Davidov and A. Rappoport, "ICWSM – A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews," in Proc. "4th International Conf. on Weblogs and Social Media" (ICWSM), pp. 162-169, May, 2010.
  - [18]. Tumasjan, Andranik, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," in Proc. "4th International Conf. on Weblogs and Social Media" (ICWSM), 2010.
  - [19]. Janyce Wiebe "Learning Subjective Adjectives from Corpora," in Proc. "17th National Conf. on Artificial Intelligence" and "12th Conf. on Innovative Applications of Artificial Intelligence", pp. 735-740, 2000.
  - [20]. S. Yang, V. Tech, A. L. Kavanaugh, and V. Tech, "Half-Day Tutorial: Collecting , Analyzing and Visualizing Tweets using Open Source Tools," in Proc. "12th International Conf. on Digital Government Research", pp. 374–375, 2011.