

Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data

Kartik Singhal, Basant Agrawal and Namita Mittal

Abstract Twitter is a microblogging website where users read and write short messages on various topics every day. Political analysis using social media is getting attention of many researchers to understand the public opinion and trend especially during election time. In this paper, we propose a novel approach based on semantics and context aware rules to detect the public opinion and further predict election results. We crawled the political tweets during the general election in India, and further evaluate our proposed approach against the election results. Experimental results show the effectiveness of the proposed rules in determining the sentiment of the political tweets.

Keywords Sentiment analysis · Social media · Political sentiment

1 Introduction

Sentiment analysis is a field of natural language processing which focuses on extraction of objective and subjective information from a natural language sentence. With the boom of Online community people are expressing their likes and dislikes towards different subjects in blogs, microblogs and social networking sites like Twitter and Facebook. Analyzing these expressions of short colloquial text [1] can yield vast information about the behavior of the people that can be helpful in many

K. Singhal (✉)

The LNM Institute of Information Technology, Jaipur, India
e-mail: gkkartiks@gmail.com

B. Agrawal · N. Mittal

Malaviya National Institute of Technology, Jaipur, India
e-mail: thebasant@gmail.com

N. Mittal

e-mail: nmittal.cse@mmit.ac.in

© Springer India 2015

J.K. Mandal et al. (eds.), *Information Systems Design and Intelligent Applications*,
Advances in Intelligent Systems and Computing 339,
DOI 10.1007/978-81-322-2250-7_46

other subjects like Political Science, opinion extraction and Human Computer Interaction (HCI).

With the ever emerging social media, more and more people are expressing their sentiments about current affairs on blogs, microblogs and social networking sites [2]. During the Indian general elections 2014, in the timeframe of 4 months, conversations regarding Indian elections were more than twice the conversations during whole of the 2013 and Indian twitter users also more than doubled.

Early research in this area focuses on adjectives and single word phrases to evaluate sentiments of the sentence like in [3]. An adjective with positive connotation can imply the overall sentiment for the subject positive and an adjective with negative connotation can imply the overall sentiment for the subject negative. However recent studies has showed that verbs and adverbs [4] and two word phrases [5] also contribute significantly to the overall opinion of the sentence. The whole process is depended upon usage of dictionary of words with their ranking (or polarity scores). This has been suggested in [6]. This method is also called lexicon based sentiment analysis. In this a pre-evaluated knowledge base consisting of words and their polarity scores are used like SentiWordNet [7] to determine relation of word phrases and there sentiment score based on classification into positivity and negativity of the subject that signifies the altitude of the author on that particular subject. Recent researches have shifted their focus on Rule Based sentiment analysis [8], Machine Learning approaches [9] and semantic meaning of the sentences [10]. Rule based techniques focuses on set of pre-defined rules which are when detected in a natural language sentence gives a definite output.

In this paper, we proposed an approach for detecting sentiment in political tweets based on our semantic rules. Proposed political sentiment analysis model is unsupervised which don't require any prior training dataset.

This paper organized as follows. Section 2 describes the related works. Section 3 discusses the proposed approach. Section 4 presents the results of the proposed approach with the discussion. Finally, Sect. 5 presents the conclusion.

2 Related Work

Previous studies [11–13] show that analyzing these sentiments and patterns can generate useful results which can be handy in determining opinions of public on elections and policies of the government. In [11], authors extract sentiments (positive, negative) as well as emotions (anger, sadness etc.) regarding the major leading party candidates and on the basis of that they calculate a distance measure. The distance measure shows the proximity of the political parties, smaller the distance higher the chances of close political connections between that parties [12] and [13] also shows how twitter data can be helpful in predicting election polls and deriving useful information about public opinions.

Existing problems in analyzing political tweets have been discussed in [14]. Sarcasm tends to reduce the accuracy of the classifier [15] shows how Sarcastic tweets in which a positive sentiment followed by an negative situation is handled. For deep analysis of the sentences, dependency parsing tool should be used which can extract relations among the words that are forming the sentence [16, 17] show the usage of Stanford Dependency Parser [18] (abbreviated from now on as SDP) in extracting these relations.

We also used categorization specified in [16] but modified them a little to suite our approach. Our categorization consists of six entities namely: Modifiers, Intensifiers, Dividers, Negations, Verbs and Objects. We believe that these entities are important as they can significantly affect sentiment of the overall sentences.

3 Proposed Approach

In this paper we proposed an approach for sentiment analysis of tweets. We believed in a common system which will be able to solve different problems like Sarcasm, Conjunction and Implicit negation combined. For this we proposed an unsupervised hybrid approach of Lexicon Based and Rule Based Sentiment Analysis which will analyze words related to other words, thus giving overall sentiment of the sentence. For lexicon, SentiWordNet is used which can give us the sentiment scores of a word. A negative score signifies negative connotation and a positive score signifies positive connotation of the word. Tweets were manually downloaded from a time period of 28 February 2014 to 28 March 2014. Our system follows in mainly 4 steps which are explained below.

3.1 *Dependency Extraction*

We used SDP to extract rules from the tweets. The sole reason is to remove extra words that are not related to overall sentiment or contribute very less to the overall sentiment. From these rules, those that are containing verbs, adjectives, adverbs, nouns, conjunctions and negations are extracted and rest are discarded.

When analyzing twitter sentences we found out that due to wrong grammatical formations, efficiency of SDP decreases which will affect our system. When SDP is unable to detect relation between two words, it uses rule 'dep' which shows unknown dependency between those words. To improve this we used Ark twitter POS tagger (abbreviated from now on as ATP) [19]. ATP enables us to determine the part of speech of the two words thus giving us the dependency.

3.2 Set Distribution

We approach the problem in set wise manner. It is easier to deal with the problem when it is divided into sets. A natural language sentence is divided into 6 sets according to their part of speech and the polarity of the whole sentence is described by describing polarity of each set in relation to the other sets. Word phrases in the sets contains reference to the words of the previous sets to which they are specifically connected. This helps us in extracting features that will be vital in classifying the sentences according to the rules in next section. The functionality of each set is explained below with the help of example sentence (1).

1. ‘BJP will make good government but still will not remove corruption from India.’
 - Set W0 (Keyword Set)—Includes Subject or Objects containing Keywords Like ‘BJP’. These contain Noun or Noun + Noun. From the above sentence (1) this set will include ‘BJP’.
 - Set W1 (Verb Set)—Includes verbs which describes the action performed by the contents of Set W0 with a reference to the specific noun to which it is connected. From the above example (1), this set includes ‘make’, ‘remove’ because of the extracted rules $nsubj(make-3, BJP-1)$ and $nsubj(remove-10, BJP-1)$ from Fig. 1 and we will extract features ‘BJP_make’, ‘BJP_remove’.
 - Set W2 (Object/subject set)—Includes objects on which the Set W0 are performing actions. This set also includes Noun and Noun + Noun. From (1) this set includes ‘government’, ‘India’ because of the relations $dojb(make-3, government-5)$, $dojb(remove-10, corruption-11)$, $prep_from(remove-10, India-13)$. We will extract features ‘make_government’, ‘remove_corruption’, ‘remove_India’.
 - Set W3 (Modifier Set)—Includes adjectival and adverbial modifiers that are providing or modifying sentiments from the above sets (W0, W1, W2). From (1) this set includes ‘good’ due to the relation $amod(government-5, good-4)$. We will extract features from this as ‘government_good’.

Fig. 1 Algorithm for sentiment score calculation

```

Function sentiscore(set(sets)) Begin:
If set is W1(verb Set) then  $S_{new} = 0.0 + S_{set}$ ;
If set is W2(object set) or W3(modifier set) then
  If  $Spre \neq 0$  then  $S_{new} = Spre * \left( \frac{S_{set}}{|S_{set}|} + S_{set} \right)$ ;
  Else  $S_{new} = S_{set}$ ;
If set is W4(intensifier set) then
  If  $Spre \neq 0$  then  $S_{new} = Spre * (1 + S_{set})$ ;
If set is W6(negator set) then  $S_{new} = -S_{prev}$ ;
Return  $S_{new}$ ;
  
```

- Set W4 (Intensifier Set)—Includes adverbial intensifiers that are strengthening or weakening the sentiment scores from the Sets above. From (1) this set includes ‘still’ from the relation *advmod(remove-10, still-7)*. We will extract feature ‘remove_still.’
- Set W5 (Buffer or Divider Set)—Includes conjunctions like ‘but’ and ‘and’ with references to two words which it is dividing. From (1) this set will include ‘but’ from the relation *conj_but(make-3, remove-10)*. We will extract feature ‘remove_make_but’.
- Set W6 (Negation Set)—includes the negation words like ‘not’, ‘never’ which flips the sentiment score from the sets above. From above example (1) this will include ‘not’ because of the relation *neg(remove-10, not-9)*. We will extract feature ‘remove_not’.

3.3 Context Rules Formation

We developed rules to determine the sentiment of tweets into positive and negative. These rules are presented in Table 1 and each rule is explained with example further. Polarity of the words is determined with the SentiWordNet. We used following abbreviations for the rules.

Example: Consider the tweet ‘AAP bhakts r always right, BJP waste time for dharnas. If u don’t trust then see it’ for the above rule. Here keyword is ‘BJP’ and set W1 includes ‘trust’ which is a positive verb in SentiWordNet and W2 includes ‘time’ and ‘waste’ both of these are minor positive and neutral nouns respectively. Notice that negator ‘don’t’ (placed in W6) is attached to trust i.e. we extract feature ‘trust_don’t’ which will reverse the polarity of Set W1 containing ‘trust’, thus classifying in Rule 1.

Table 1 Context rules

Rules	Set W1 Verb set	Set W2 Object/Subject set	Set W3 Adjective set	Set W4 Adverb set	Polarity
1	VB-	N+/neutral	*	*	-ve
2	VB-	N-	J-	*	+ve
3	VB+/neutral	N-	*	*	-ve
4	VB+/neutral	N+/neutral	J+/neutral	*	+ve
5	VB+/neutral	N+/neutral	J-	*	-ve
6	VB+/neutral	N+/neutral	J+	RB+	+ve
7	VB+/neutral	N+/neutral	J-	RB+	-ve
8	VB+/neutral	N+/neutral	J+	RB-	+ve
9	VB+/neutral	N+/neutral	J-	RB-	-ve

VB verb, N noun, J adjective, RB adverb, * doesn’t matter, + positive, - negative

3.4 Determining Sentiment Scores

Once the rule formation occurs, sentiment scores are calculated using SentiWord-Net. We used method similar to specified in [16] in calculating and distributing scores. Let *Spre* be the polarity from the previous sets with which it is connected to, *Sset* be the polarity of particular set and *Snew* be the updated polarity. Figure 1 shows the algorithm used for the sentiment score calculation.

4 Results

We prepared a datasets of total 259 tweets from the date of 28 February 2014 to 28 May 2014 just before the time of Indian elections to know the public trend and general opinion about the elections. Among the total tweets 116 are positive, 92 are negative and remaining 51 are objective tweets. We used accuracy as evaluation measure and it is computed by dividing the correctly classified tweets with total number of tweets. Our approach correctly predicted 76 positive tweets and 55 negative tweets. Further, we investigated manually that tweets containing colloquial language (containing Hindi words) is 56 out of which 20 were positive and 17 were negative. We removed these tweets from total tweets. Results are presented in Table 2.

Next, we tried to model elections in National Capital Territory (NCT) Delhi region. For this, we manually downloaded 106 tweets giving sentiments for the Aam Admi Party (AAP) and its party leader Arvind Kejriwal from the same time period by the users of Delhi Region. We investigated tweets with #AamAdmiParty, #AAP and #ArvindKejriwal. Results are presented in Table 3.

Table 2 Results

Accuracy—positive tweets	$(76/96) \times 100 = 79.17 \%$
Accuracy—negative tweets	$(55/75) \times 100 = 73.33 \%$
Overall accuracy	$(131/171) \times 100 = 76.61 \%$

Table 3 Results related to modelling elections in NCT Delhi

Number of tweets evaluated containing keyword AAP	106
Tweets containing positive sentiment towards AAP	37
Percentage of users positive about AAP	$(37/106) \times 100 = 34.91 \%$
Tweets containing negative sentiment towards AAP	51
Percentage of users negative about AAP	$(51/106) \times 100 = 48.11 \%$

4.1 Discussions and Comparison with Other Approaches

Above results shows that 34.91 % of users were positive towards AAP party in NCT Delhi region. From the Indian General Election results 2014 we know that all the 7 seats of Delhi region were won by Bhartiya Janta Party (BJP). Although AAP was not able to win any seat in NCT Delhi, there voting share in the elections were 32.90 % Ref. [20]. This gives us an error percentage of 6.11 %. So we were able to predict the voting share of AAP with acceptable error percentage.

We compared our proposed approach with the state-of-art approaches. Table 4 presents some cases of where other approaches fails whereas proposed approach performs better than other methods. The example tweets are chosen from the dataset according to the classification type by algorithm.

Table 4 Handling problems in existing strategies

Type	Example tweets	Di Caro and Grella [16]	Riloff et al. [15]	Blenn et al. [1]	Tan et al. [17]	Our approach
Noun driven sentiment	If AAP comes to power, they will form worst dictators	✗	✗	✗	✓	✓
Verb driven sentiment	AAP bhakts r always right, BJP waste time for dharnas. If u don't trust then see it	✓	✓	✓	✓	✓
Adjective driven sentiment	AAP is worst in forming and managing policies	✓	✓	✓	✓	✓
Adverb driven sentiment	Kejriwal is somewhat less insane then his oppositions	✓	✓	✗	✓	✓
Conjunctions	The IT cell of AAP may be good in photoshop but lack brains in logic when they create crowd	•	•	•	•	✓
Explicit negation	AAP bhakts r always right, BJP waste time for dharnas. If u don't trust then see it	✓	✓	✓	✓	✓
Implicit negation	Meanwhile, Arvind Kejriwal opens AAP's doors to most wanted Maobadi terrorist	✗	•	✗	✗	•
Sarcasm	Wow!!! AAP couldn't be more right in forming the policies	✗	•	✗	✗	•
Sarcasm (positive negative situation)	Absolutely adore it when my bus is late	✗	✓	✗	✗	✓

✓ handled, ✗ not handled, • handled in some cases

5 Conclusion

People are increasingly using Social media to express their opinion. And, Twitter is a great source to investigate the public opinion especially during election time. Observing the results has led us to believe that there is a great scope in analyzing Indian political twitter data and considering its sentiment alone can result in giving a general idea about the election results. In this paper, we proposed various rules based on semantic structure of the sentence. Experimental results show the effectiveness of the proposed approach over existing methods.

References

1. Blenn, N., Charalampidou, K., Doerr, C.: Context-sensitive sentiment classification of short colloquial text. In: Proceedings of IFIP'12, pp. 97–108, Prague, Czech Republic (2012)
2. Mittal, N., Agarwal, B., Agarwal, S., Agarwal, S., Gupta, P.: A hybrid approach for twitter sentiment analysis. In: 10th International Conference on Natural Language Processing (ICON), pp. 116–120 (2013)
3. Agarwal, B., Mittal, N.: Prominent feature extraction for review analysis: an empirical study. *J. Exp. Theor. Artif. Intell.* (2014). doi:[10.1080/0952813X.2014.977830](https://doi.org/10.1080/0952813X.2014.977830)
4. Subrahmanian, V.S., Reforgiato, D.: Ava: adjective-verb-adverb combinations for sentiment analysis. *Intell. Syst.* **23**(4), 43–50 (2008)
5. Turney, P.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of 40th Meeting of the Association for Computational Linguistics, pp. 417–424, Philadelphia, PA (2002)
6. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. *Comput. Linguis.* **37**(2), 267–307 (2011)
7. Esuli, A., Sebastiani, F.: SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of 5th International Conference on Language Resources and Evaluation (LREC), pp. 417–422 (2006)
8. Romanyshyn, M.: Rule-based sentiment analysis of ukrainian reviews. *Int. J. Artif. Intell. Appl.* **4**(4), 103–111 (2013)
9. Kessler, J.S., Nicolov, N.: Targeting sentiment expressions through supervised ranking of linguistic configurations. In: 3rd International AAAI Conference on Weblogs and Social Media (2009)
10. Bandyopadhyay, S., Mallick, K.: A new path based hybrid measure for gene ontology similarity. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **11**(1), 116–127 (2014). doi:[10.1109/TCBB.2013.149](https://doi.org/10.1109/TCBB.2013.149)
11. Tumasjan, A., Sprenger, T.O., Sandner, P., Welpe, I.: Predicting elections with twitter: what 140 characters reveal about political sentiment. In: Proceedings of ICWSM (2010)
12. O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: linking text sentiment to public opinion time series. In: ICWSM (2010)
13. Birmingham, A., Smeaton, A.F.: On using twitter to monitor political sentiment and predict election results. In: Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011-IJCNLP), pp. 2–10, Chiang Mai, Thailand (2011)
14. Bakliwal, A., Foster, J., Pui, J.V.D., O'Brien, R., Tounsi, L., Hughes, M.: Sentiment analysis of political tweets: towards an accurate classifier. In: Proceedings of NAACL Workshop on Language Analysis in Social Media, pp. 49–58 (2011)

15. Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., Huang, R.: Sarcasm as contrast between a positive sentiment and negative situation. In: EMNLP 2013, pp. 704–714 (2013)
16. Di Caro, L., Grella, M.: Sentiment analysis via dependency parsing. *Comput. Stan. Interfaces* (2012)
17. Tan, L.K.W., Na, J.C., Theng, Y.L., Chang, K.: Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. *J. Comput. Sci. Technol.* **27**(3), 650–666 (2012)
18. De Marneffe, M., MacCartney, B., Manning, C.: Generating typed dependency parse from phrase structure parses. *LREC* (2006)
19. Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., Smith, N.A.: Part-of-speech tagging for twitter: annotation, features, and experiments. In: *Proceedings of ACL* (2011)
20. http://eci.nic.in/eci_main1/GE2014/PC_WISE_TURNOUT.htm