

Affect Analysis of Multilingual Tweets for Predicting Voting Behavior



Lata Gohil, Dharmendra Patel

Abstract: Social media has been proved as wild card for its role in election campaign across the globe. It has been used for general election of India in year 2014 and year 2019 by political parties for election campaign. Thus social media provides opportunity for electoral prediction. Users from India use regional languages in addition to English language on social media. Multilingual data likely to give better prediction compared to single language data. Affect analysis gives deeper insight compared to sentiment analysis. This research study aims to predict voting behavior for 2019 general election of India using affect analysis of multilingual tweets. Three languages namely English, Hindi and Gujarati are used for this study. Volume-based method and machine learning algorithm based method are two approaches widely used in literature for electoral prediction. In this research study hybrid approach is used along with consideration of ratio of positive count and negative count of tweets. Experiment result shows efficacy of the proposed approach.

Keywords : Multilingual, Social Media, Twitter, Emotion Analysis, Sentiment Analysis, Opinion Mining, Election

I. INTRODUCTION

In India's general election 2019, social media has been used extensively by political parties for campaign. Social media provides opportunity to politician to get in touch with voters online and interact with them. Social media users express their views on discussion or campaign related to election. Thus, social media plays important role in influencing voting behavior of its users.

Twitter is one of the social media which was widely used for election campaign in 2019 Indian general election by political parties. As an endeavor to educate voters on various aspects of election and to encourage them to vote, the Election Commission of India had joined Twitter on 18th March 2019 [1] [2].

India is multilingual country. "The Constitution of India has designated the official languages of India as Hindi and English" [3] [4]. Considerable amount of citizens of India are bilingual and trilingual. Twitter supports many languages other than English including many regional languages of India.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Lata Gohil*, Computer Science and Engineering Department, Institute of Technology, Nirma University, India

Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), CHARUSAT, Changa, India

Dharmendra Patel, Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), CHARUSAT, Changa, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Decision making is considered as one of the basic cognitive process of human behaviors. "Emotions are intense feelings that are directed at someone or something" [5]. "Affect is a broad term that covers a wide range of feelings that people experience. It is a concept comprises of emotion and mood" [6]. Thus, emotions and affect play vital role in decision making.

This research study performs affect analysis of multilingual tweets to predict voting behavior. Electoral prediction related research study has been presented in section 2. Motivation for research study attempted and presented in this paper is mentioned in section 3. The methodology applied for affect analysis is explained in section 4. Experiment and result are discussed in section 5 and 6 respectively. Section 6 provides conclusion with future scope for enhancement of this research study.

II. RELATED WORK

Social media has remarkable impact on voting behavior of young voters and it also influence voting decision of its highly active users [7]. The research study in [8] concludes that sentiment analysis of tweets can be used for prediction of election result.

Tweet volume and sentiment analysis are two approaches used in [9] for election result prediction and its result shows that sentiment analysis based approach can be used for prediction but tweet volume based approach is not good predictor for election result which is conflicting with result of previous study presented in [10] and [11].

Dynamic keywords and topic modeling methods are explored with sentiment and volume based method to predict electoral outcome in [12]. While study in [13] has employed analytical procedure based approach and supervised approach to predict result of election.

III. MOTIVATION

In last one decade, social media has played active role for election campaign. Social media has proved its effectiveness for election campaign in 2014 general election of India. Twitter supports regional languages of India and users from India use their regional languages in addition to official languages of India for expressing their views and posting comments. As substantial amount of regional languages tweets gets generated by twitter's users, mining of one language tweets is not sufficient and thus extracting knowledge from multilingual tweets becomes an important task.



Affect Analysis of Multilingual Tweets for Predicting Voting Behavior

Considerable amount of research studies have been done for predicting voting behavior by performing sentiment analysis on tweets of only single language. Regional languages of India are very less explored for this task. Affect analysis provides more insight compared to sentiment analysis as later does analysis at coarse-level while former does analysis at fine-level. This scenario motivates us to perform affect analysis of multilingual tweets to predict voting behavior. Three languages namely Gujarati, Hindi and English are considered for this study. As far as our knowledge is concerned, this research study is the first attempt to perform multilingual affect analysis in context of Indian languages for predicting voting behavior.

IV. METHODOLOGY

Study on emotion analysis of multilingual tweets has been attempted in [14]. This paper extends the study of [14] for performing affect analysis of multilingual tweets. Result of study presented in [14] shows that performance of hybrid approach CS-SN-ML when applied on majority votes data set with LinearSVC machine learning algorithm is higher compared to other approaches used. Hence, we use the architecture presented in [14] along with hybrid approach where LinearSVC is used with our proposed feature generation algorithm. Affect comprises both emotions and mood [6]. Feature generation algorithm is proposed to generate features which takes both emotion and mood in consideration. We gave CS-SN-Affect name to this algorithm. Primary features and secondary features are prepared using TF-IDF weighting and SenticNet [15] resource respectively.

A. CS-SN-Affect Algorithm

The sentic vectors of primary features, discrete eight emotions and their primary mood tag and secondary mood tag from SenticNet [15] resource are considered for generation of secondary features. Eight emotions namely anticipation, joy, disgust, trust, sadness, surprise, anger, fear specified in Plutchik's wheel of emotions [16] are taken into consideration.

Algorithm: Pseudocode for generating features taking into consideration both emotions and mood.

Input: Cleaned Tweets prepared through preprocessing

Output: Vector of secondary features

1. Prepare primary features from inputted cleaned tweets using TF-IDF weighting

$$PFS = \{pf_1, pf_2 \dots pf_n\}$$
2. $\forall pf_i \in PFS$, if it is in Gujarati language then convert it in Hindi by performing translation
3. Prepare set of pfs_i where pfs_i is sentic vector of pf_i

$$PSV = \{pfs_1, pfs_2 \dots pfs_n\}$$
4. Prepare set of $pfspm_i$ where $pfspm_i$ is sentic vector of primary mood tag of pf_i

$$PSVPM = \{pfspm_1, pfspm_2 \dots pfspm_n\}$$
5. Prepare set of $pfssm_i$ where $pfssm_i$ is sentic vector of secondary mood tag of pf_i

$$PSVSM = \{pfssm_1, pfssm_2 \dots pfssm_n\}$$

6. Prepare set of sentic vector for each emotion e_i where $e_i \in E$ and $E = \{e_1, e_2, \dots e_8\}$

$$ESV = \{es_1, es_2 \dots es_8\}$$
7. Prepare set of epm_i where epm_i is sentic vector of primary mood tag of e_i

$$ESVPM = \{epm_1, epm_2 \dots epm_8\}$$
8. Prepare set of esm_i where esm_i is secondary mood tag of e_i

$$ESVSM = \{esm_1, esm_2 \dots esm_8\}$$
9. Prepare eight secondary feature sf_{ij} for each primary feature pf_i by taking each $e_i \in E$ into consideration.

$$SF = \{sf_{11}, sf_{12}, sf_{13} \dots sf_{18},$$

$$sf_{21}, sf_{22}, sf_{23} \dots sf_{28},$$

$$\dots \dots \dots$$

$$sf_{n1}, sf_{n2}, sf_{n3} \dots sf_{n8}\}$$

where $sf_{ij} = \text{cosine_similarity}(pfs_i, es_j)$

$$+ \text{cosine_similarity}(pfspm_i, epm_j)$$

$$+ \text{cosine_similarity}(pfssm_i, esm_j)$$

V. EXPERIMENT AND RESULT

Two experiments are performed. First experiment is for assessing performance of proposed feature generation algorithm. Second experiment is for performing affect analysis to predict voting behavior.

A. Performance of CS-SN-Affect Algorithm

Secondary features are generated using CS-SN-Affect algorithm and then used by machine learning algorithm LinearSVC. Majority data set is used given in [14] for performing this experiment. F-measure is used as performance measure. Results are presented in Table I.

Table I: Average F-measure for Majority Votes Dataset using LinearSVC

| Language | CS-SN | CS-SN-Affect |
|----------|-------|--------------|
| English | 0.84 | 0.84 |
| Hindi | 0.78 | 0.77 |
| Gujarati | 0.78 | 0.77 |

Average F-measure presented in Table I shows that there is no significant difference in result of CS-SN algorithm and CS-SN-Affect algorithm. "Moods are feelings that tend to be less intense compared to emotions and often having absence of contextual stimulus" [17]. Cause for Mood is often general and unclear. This may be the reason for insignificant difference of performance.

B. Affect analysis for predicting voting behavior

Two data sets for leading political parties namely BJP and Congress are prepared from majority votes data set used in [14]. Data set for BJP and Congress are prepared using tweets containing specific hashtags.

The purpose of this experiment is to perform affect analysis of multilingual tweets for predicting voting behavior. Social media users' affection towards political parties is used to determine the voting behavior and aggregation of the same is used for predicting electoral result.

Features are generated using CS-SN-Affect algorithm and machine learning algorithm LinearSVC is used for multilabel classification of multilingual tweets. Tweets of three languages namely English, Hindi and Gujarati are taken into consideration.

For each emotion categories, positive count and negative count are taken into consideration. Positive count (PC) is count of tweets which are correctly classified and are representative of respective emotion. Negative count (NC) is count of tweets which are correctly classified and are not representative of respective emotion. Ratio of PC and NC is taken and presented in Table II.

Table II: Emotion wise Ratio Measure

| Emotion | BJP | Congress |
|--------------|------|----------|
| Anger | 0.74 | 0.73 |
| Anticipation | 1.50 | 1.33 |
| Disgust | 0.79 | 0.79 |
| Fear | 0.83 | 0.84 |
| Joy | 0.84 | 0.86 |
| Sadness | 0.71 | 0.67 |
| Surprise | 0.89 | 0.75 |
| Trust | 0.81 | 0.85 |

Aggregation of ratio is taken to understand overall positive emotion and negative emotion. Average of ratio of positive emotions and average of ratio of negative emotions are presented in

Table III.

Table III: Ratio for Positive and Negative Emotions

| Emotion | BJP | Congress |
|-------------------|------|----------|
| Positive Emotions | 0.94 | 0.91 |
| Negative Emotions | 0.77 | 0.76 |
| Difference | 0.17 | 0.15 |

Difference of ratio of positive emotion and negative emotion represents inclination of voters for respective political parties. Though the prediction is correct about winning party, there is no significance difference in overall emotion for both parties and this may be due to only three languages are taken into consideration. India is multilingual country and "The 2001 Census recorded 30 languages which were spoken by more than a million native speakers and 122 which were spoken by more than 10,000 people" [18] [19]. Thus, taking tweets of more number of languages into consideration likely to improve the prediction. Tweets were collected only for few days ago of election. Collection of tweets for larger span before election may also improve the prediction.

VI. CONCLUSION AND FUTURE WORK

This study aims to predict voting behavior using affect analysis of tweets of English, Hindi and Gujarati languages. Feature generation algorithm is proposed which takes both

emotions and mood into consideration. LinearSVC algorithm and ratio of positive count and negative count of tweets are used for prediction. Two political parties of India namely BJP and Congress are taken into consideration for this study. Result of prediction about winning party is correct but there is scope of further improvement. More number of regional languages tweets and collection of tweets for larger number days before election are likely to improve the prediction result.

REFERENCES

1. "Election Commission of India #SVEEP (@ECISVEEP) | Twitter." [Online]. Available: <https://twitter.com/ecisveep?lang=en/>. [Accessed: 19-Nov-2019].
2. "Election Commission of India joins Twitter ahead of Lok Sabha Elections 2019 - Technology News." [Online]. Available: <https://www.indiatoday.in/technology/news/story/election-commission-of-india-joins-twitter-ahead-of-lok-sabha-elections-2019-1484627-2019-03-23>. [Accessed: 19-Nov-2019].
3. "Profile - The Union - Official Language - Know India: National Portal of India." [Online]. Available: <https://knowindia.gov.in/profile/the-union/official-language.php/>. [Accessed: 19-Nov-2019].
4. "Multilingualism in India - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Multilingualism_in_India. [Accessed: 19-Nov-2019].
5. N. H. Frijda, "Moods, emotion episodes, and emotions.," 1993.
6. P. Ekkekakis, "Affect, mood, and emotion," *Meas. Sport Exerc. Psychol.*, vol. 321, 2012.
7. A. Biswas, N. Ingle, and M. Roy, "Influence of social media on voting behavior," *J. Power, Polit. Gov.*, vol. 2, no. 2, pp. 127–155, 2014.
8. M. Ramzan, S. Mehta, and E. Annapoorna, "Are tweets the real estimators of election results?," in *2017 Tenth International Conference on Contemporary Computing (IC3)*, 2017, pp. 1–4.
9. B. Heredia, J. Prusa, and T. Khoshgoftaar, "Exploring the effectiveness of twitter at polling the united states 2016 presidential election," in *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, 2017, pp. 283–290.
10. A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Fourth international AAAI conference on weblogs and social media*, 2010.
11. A. Boutet, H. Kim, and E. Yoneki, "What's in your tweets? I know who you supported in the UK 2010 general election," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
12. V. K. Jain and S. Kumar, "Towards Prediction of Election Outcomes Using Social Media," *Int. J. Intell. Syst. Appl.*, vol. 9, no. 12, p. 20, 2017.
13. M. Maldonado and V. Sierra, "Twitter Predicting the 2012 US Presidential Election?: Lessons Learned from an Unconscious Value Co-Creation Platform," *J. Organ. End User Comput.*, vol. 28, no. 3, pp. 10–30, 2016.
14. L. Gohil and D. Patel, "Multilabel classification for emotion analysis of multilingual tweets," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 4453–4457, 2019.
15. E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
16. R. Plutchik, "Emotion: A psychoevolutionary synthesis: Harpercollins College Division," 1980.
17. D. Hume, "Emotions and moods," *Organ. Behav.*, pp. 258–297, 2012.
18. "Languages of India - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Languages_of_India. [Accessed: 01-Dec-2019].
19. "Census of India: General Note." [Online]. Available: http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/gen_note.html. [Accessed: 01-Dec-2019].



Affect Analysis of Multilingual Tweets for Predicting Voting Behavior

AUTHORS PROFILE



Lata Gohil is working as Assistant Professor in Computer Science and Engineering Department, Institute of Technology, Nirma University. She has received MCA degree from Gujarat Vidyapith, Ahmedabad. She has qualified GSET and GATE. Her research area is Information Retrieval and Text Mining. She is pursuing PhD from CHARUSAT.



Dr Dharmendra Patel received his Master of Computer Application degree from North Gujarat University. He received his PhD degree in computer science from Kadi Sarva Viswavidyalaya. His area of research is Web Mining, Fog Computing, Image Processing, Internet of Things etc. He has published 20 papers in national/international journal of repute. Currently he is working as an associate professor at CHARUSAT, Changa.