

Sentiment Analysis on Twitter using Multiple Emoji and Sarcasm Detection

H. C. Jadav¹ K. U. Khimani² D.R. Chauhan³

^{1,2}Department of Information Technology

^{1,2}VVP Engineering College Rajkot, Gujarat, India ³Christ Polytechnic Institute Rajkot, Gujarat, INDIA

Abstract— Millions of users share opinions on different aspects of life every day on social media. User can give their gist regarding the recent events in their surrounding and give suggestions to improve surroundings in text-based format while conveying their emotions which they are not able to easily verbalize using emoticons and Emoji. Spurred by that growth, companies and media organizations are increasingly seeking way to mine information. For better understanding of people's opinion, it is important to analyze different semiotics as well as sentence. We propose a method Multiple Emoji Identification and Sarcasm Detection (MEISD) to provide an approach to determine sentiment score of a tweet with semiotics with multi-dimensional sentiment analysis and sarcasm detection. We have created semiotic dictionary which have sentiment score for each semiotic with sentiment expressed by it most frequently. Also, we identify multiple emoji on tweet that will helpful in sentiment analysis and work when user use multiple emoji on single tweet and find tweet that contain sarcasm or not, also classify given tweet.

Keywords: Sentiment Analysis, Sarcasm Detection, Sarcasm, Emoticon

I. INTRODUCTION

The purpose of social media websites like Twitter, Tumbler, and Facebook is that its user can express their feelings without being pressurized by anyone. User can give their point of view regarding the recent events in their surroundings as well as give suggestions to improve surroundings in text-based format while conveying their emotions which they are not able to easily verbalize using emoticons and emoji. For better understanding of people's opinion, it is important to analyze this semiotics as well as sentence[1].

In the epoch of social network, people's opinion has become one of the extremely important sources for various services. Opinion on the Internet could be in millions. Sentiment analysis is one of the major focuses in the field of research to extract the immanent data to trail and discern customer opinions [2].

Twitter has data of public opinions of half a billion people on planet, whether be it an ordinary working class man, a politician or any global super star. In fact, companies manufacturing different products are using these micro-blogs to obtain a gist of common opinion about their product. Political parties are interested to know general reaction of the citizens regarding their campaigns. Social organizations may want to have information about peoples view on present discussions [3].

The challenges involve in sentiment analysis is the presence of sarcastic sentiment. Due to this, most of the existing systems for sentiment analysis fails in detecting the actual sentiment value. Sarcasm is a particular type of sentiment that macmillan dictionary, sarcasm is known as

“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”[4].

A sarcasm detection system is important for application like sentiment analyzer, review processing system and natural language processing systems. Sarcasm is one of the major linguistic concept used in social media nowadays. People use sarcastic comments and messages in order to funny or to irritate and hurt others [6].

The detailed analysis of the datasets used for sarcasm detection shows that either one of them is presence in more than 70% of tweets [6].

II. RELATED WORK

Darsha R. Kamal et al. [1]. Author has proposed a system that is considering twitter data for multidimensional sentiment analysis. They retrieve Tweet through API (input). Tweet preprocessing is required to remove noisy data like URL and @user mentioned in tweet is not necessary for sentiment analysis. For sentiment score calculation they split tweet into two string: Text String and Semiotic string. For calculating text sentiment score they consider average sentiment score of word in the text. For semiotic sentiment score they consider semiotic dictionary which also have the sentiment expressed by the semiotic [1]. This research work does not work in a tweet that contain multiple emoji or different emoji and this is also not work in sarcastic tweet this is the main research gap of this paper.

Snehal Kale Vijaya Padmadas [2]. Proposed the system in which they extract the tweets information from Twitter using web mining techniques and analyze these tweets using sentiment analysis. Sentiment analysis or web mining can be defined as the process of extracting knowledge from sentiments or tweets of users about a topic or problem. They identify tweets in a large unstructured or structured data then the data collected would be processed to remove useless content. Then they analyze their polarity, and will categorize them into different categories basis the Naïve Bayes and Maximum Entropy, and Negation. In this, they using semantic analysis and classification. Machine learning will be used to analyze the tweets and to tag obtained tweet in a predefined category. The results can be used for various purposes such as guiding decisions to improve the system based on sentiment classification [2]. They compare accuracy of Naïve Bayes and Maximum Entropy score for text polarity but there is no results shown in research paper of sarcastic tweet and tweet that contain multiple or different emoji.

Srinidhi Bhat, Saksham Garg and Poornalatha G [3]. They enter the word of interest, which then searched all over Twitter based on the occurrence of that word in a particular tweet. Using Twitter API they scrape tweets, which was extensively used in data extraction process. Next, they make a dictionary for positive and negative words and assign

priorities as: Positive Words (single occurrence at that instant): 1 Negative Words (Single Occurrence at that instant): -1. In this approach, in case a word appears more than once next to each other, they assign them a higher score based on its polarity as the sentiment conveyed in such a case would be different. [3]. This model used could not detect and identify sarcastic statements.

Shubham Rendalkar, Chaitali Chandankhede [11]. According to Author study the sarcasm identification model is a novel approach based on emotion model. The sarcasm identification model uses different algorithms, libraries and methods in emotion detection phase and its result is used for sarcasm detection. So it is too much dependent on the emotion identification module which poses risk at times. Most of the time Facebook comments are consists of words, hashtags and the emoticons. The system also considers hash tags and emoticons for sarcasm detection which are an important feature set of Facebook posts. They combine word based detection and emoticons based detection approach to find sarcasm.

Chae Won Park and Dae Ryong Seo [9]. author experimental procedure consisted of collecting Twitter corpus and classifying using the lexicon. After classification, the proportion of positive, negative, and neutral opinions was extracted from the lexicon [9].

III. PROPOSED TECHNIQUE

We propose a method to provide an approach to determine sentiment score of a tweet with semiotics with multi-dimensional sentiment analysis and sarcasm detection. We have created semiotic dictionary which have sentiment score for each semiotic with sentiment. Also, we identify multiple or different emoji on tweet that will helpful in sentiment analysis and work when user use multiple or different emoji on single tweet and find tweet that contain sarcasm or not, also classify given tweet.

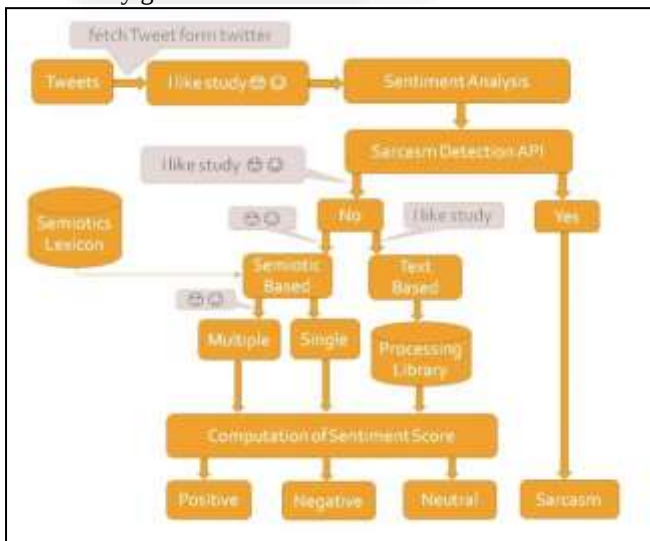


Fig. 3.1: Proposed System

A. Proposed System:

In this study, we have to retrieve Tweets through API (input). Twitter provides different API. Streaming API is used for gathering live tweets from twitter. Using this API we can

collect huge amount of tweets. With help of this API we can create feed from user stream or any particular site stream.

First of all, we have to fetch tweets from Twitter. For that we are use Php twitter text formatter which is Twitter Streaming Library.

For sarcasm detection we use ParallelDot API. ParallelDots Sarcasm Detection API maintains high accuracy in the real world, and is robust against tricky sentences with contradictory sarcasm like - "I like long walks (positive), especially when they are taken by people who annoy me (negative)". This Sarcasm Detection API accurately identifies presence of sarcasm in the text content incorporated from sources like twitter, Blogs, Articles, forums, consumer reviews, surveys etc. This API internally process and call its function for sarcastic sentence.

If text is sarcastic then it display result text is sarcastic. Otherwise it calls PHP Insight Library. PhpInsight is a Naive Bayes classifier to calculate sentiment. The program uses a database of words categorized as positive, negative or neutral. This library consists of PHP code that will classify users live tweet and give text polarity as positive, negative or neutral.

This library consists of more than 9000 positive words, 9000 negative words, 400 neutral words and more than 500 ign words. Using this dictionary our method shows the result that the given tweet is positive, negative or neutral.

B. Sentiment Score Calculation:

For sentiment score calculation we split our tweet into two strings: Text string and Semiotic string. For calculating text sentiment score, we have considered average of sentiment score of words in the text. For semiotic sentiment score we have considered semiotic dictionary which also have the sentiment score expressed by the semiotic. Then addition of text score and semiotic score will be done and display final score of sentiment analysis.

C. Proposed Algorithm:

The algorithm will calculate sentiment score for the entire tweet considering both text and semiotics.

By using methods for calculating sentiment score of text and semiotic, we have developed an algorithm which is explained below.

Input: Tweet corpus and Semiotic dictionary

Output: Polarity of text or sarcastic tweet and Sentiment score of tweet

- 1: procedure
- 2: Fetch tweets using twitter API
- 3: Sentiment Analysis
- 4: Check Sarcasm using sarcasm Detection API
- 5: If (Sarcasm = Yes)
- 6: Then
- 7: Tweet is Sarcastic
- 8: Else
- 9: TweetDetection : Two Files :- Semiotic & Text
- 10: Semiotic Sentiment Score Calculation
- 11: If (Semiotic = single)
- 12: Then
- 13: Calculate SemioticScore
- 14: Else
- 15: Semiotic 1 + Semiotic N

```

16: Calculate SemioticScore
17:
18: If (SemioticScore > 0.2)
19: Then
20: SemioticPolarity = Positive
21:
22: Else If (SemioticScore < -0.2)
23: Then
24: SemioticPolarity = Negative
25:
26: Else
27: SemioticPolarity = Neutral
28:
29: Text Sentiment Score Calculation
30:
31: Processing Library for text
32: If(TextScore > 0.2)
33: Then
34: TextPolarity = Positive
35:
36: ElseIf (TextScore < -0.2)
37: Then
38: TextPolarity = Negative
39: Else
40: TextPolarity = Neutral
41:
42: Tweet sentiment score = (text sentiment score + semiotic
    sentiment score) / 2
43: Result
4. Result Analysis

```

Using our research code we had seen how sentiment analysis will be done, now we analyze the research result.

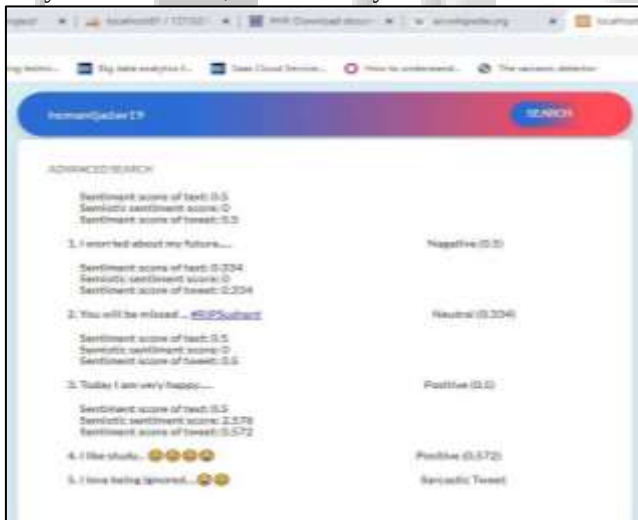


Fig. 4.1: Sentiment Score Result

Previous research work is not work on tweet that contain multiple emoji (semiotics) of given tweet and are not able to find sarcasm on given tweet, that means there is no hybrid or combine approach which work on tweet which contain more than one emoji or different emoji and sarcasm[1]. In our algorithm gives the accurate results and work on tweet that contain multiple emoji and also able to find sarcasm in a given tweet. A comparison chart is given in following figure 4.2.

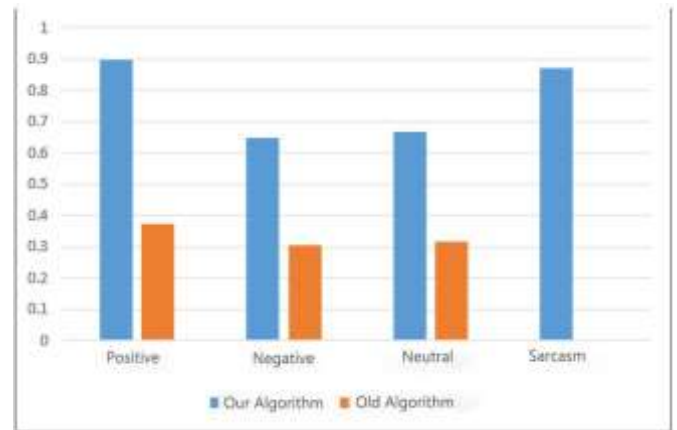


Fig. 6.14: Comparison of Our Algorithm and Old Algorithm Results with Tweets^[1]

IV. CONCLUSIONS AND FUTURE WORK

A. Conclusion:

Earlier, there is sentiment score calculation is done separately for semiotics and text and not work on multiple or different emoji of tweet. We have combined both and calculate tweet sentiment score using semiotic and text for more accurate sentiment score for the given tweet. Also, we identify multiple or different emoji on tweet that will helpful in sentiment analysis and work when user use multiple or different emoji on single tweet and find tweet that contain sarcasm or not, also classify given tweet

B. Future Work:

In our work we mostly focus on English language tweets. In future this work can be additionally reached out later by work on different languages in twitter to give more accurate result.

REFERENCES

- [1] Darsha Chauhan, Kamal Sutaria. (2018). "Multidimensional Sentiment Analysis on Twitter with Semiotics" In Proceeding of International Journal of Information Technology.
- [2] Snehal Kale, Vijay Padmadas (2017). Sentiment Analysis of Tweets Using Semantic Analysis. In Proceeding of 2017 Third International Conference on Computing, Communication, Control And Automation (ICCUBEA)
- [3] Srinidhi Bhat, Saksham Garg and Poornalatha G. (2018). Assigning Sentiment Score For Twitter Tweets. In Proceeding IEEE 2018. (PP. 934-937).
- [4] Santosh Kumar Bharti, Reddy Naidu and Korra Sathya Babu. (2017). "Hyperbolic Feature-based Sarcasm Detection in Tweets: A machine Learning Approach". In Proceeding of IEEE Conference.
- [5] Tanya Jain, Nilesh Agrawal, Garima Goyal and niyati Agrawal. (2017). "Sarcasm Detection of Tweets: A Comparative Study". In Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3), 10-12 August 2017, Noida, India
- [6] Sreelakshmi K and Rafeeqe P C. (2018). "An Effective Approach for Detection of Sarcasm in Tweet". In Proceeding of 2018 International CET Conference on

- Control, Communication, and Computing (IC4) | July 05 – 07, 2018 | Trivandrum. (pp. 377-383).
- [7] Shiha, M., & Ayvaz, S. (2017). "The Effects of Emoji in Sentiment Analysis." *International Journal of Computer and Electrical Engineering (IJCEE.)* Vol, 9(1), 360-369.
- [8] Harpreet Kaur, Veenu Mangat and Nidhi. (2017). "A Survey of Sentiment Analysis Technique". In *Proceeding of International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017)*. (pp. 921-925).
- [9] Chae Won Park and Dae Ryong Seo (2018). "Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistant". In *proceeding 2018 5th International Conference on Industrial Engineering and Applications*. (pp. 495-498).
- [10] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proc. The Eighth International AAI Conference on Weblogs and Social media*, May. 2014, pp. 2-10.
- [11] Shubham Rendslkar and Chaitali Chandankhede. (2018). "Sarcasm Detection Of Online Comments Using Emoticon Detection". In *Proceeding of International Conference on Inventive Research in Computing Applications (ICIRCA 2018) IEEE Xplore Compliant Part Number: CFP18N67-ART; ISBN: 978-1-5386-2456-2*, (pp. 1244-1249).
- [12] Dwi A. P. Rahayu, Soveatin Kuntur & Nur Hayatin. Sarcasm Detection on Indonesian Twitter Feeds. In *Proceeding of EECSI 2018, Malang - Indonesia*, 16-18 Oct 2018 (pp. 137-141)
- [13] Ha, I., Back, B., & Ahn, B. (2015). Mapreduce functions to analyze sentiment information from social big data. *International Journal of Distributed Sensor Networks*. ACM
- [14] D. Yadav, P., & Pandya, D. (2017, February). SentiReview: Sentiment analysis based on text and emoticons. In *Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on* (pp. 467-472). IEEE.
- [15] Shiha, M., & Ayvaz, S. (2017). The Effects of Emoji in Sentiment Analysis. *International Journal of Computer and Electrical Engineering (IJCEE.)* Vol, 9(1), 360-369.
- [16] Sentiment Analysis
https://en.oxforddictionaries.com/definition/sentiment_analysis
- [17] E. R. Team, "2016 emoji report." http://cdn.emogi.com/docs/reports/2016_emoji_report.pdf