

# A Review on Sentiment Analysis of Twitter Data

Aarti Chugh

Assistant Professor, Department of Computer Science and Engineering, Amity University, Haryana, India

**ABSTRACT:** In today's era, the textual data is rapidly growing all over Internet. Social networks are big platforms for many companies to get deep insight into people behavior and requirements about their products and services. Among these social platforms, Twitter is an enormously popular microblog on which clients may voice their opinions. The strings of characters called tweets carries people opinions which are widely used for sentiment analysis. This paper focuses on some of recent famous techniques used for sentiment analysis of tweets.

**KEYWORDS:** Twitter; sentiment analysis; social networks; text mining.

## I. INTRODUCTION

Sentiment refers to emotions or views of the users. Many terms can be used to represent people mood or reviews like good, bad, sad, happy, excellent, neutral etc. The analysis of these emotions is called as sentiment analysis. In other words, we can say that, it is a natural language processing task which uses computational approach that identify the opinion of user and classify them into negative, positive or neutral. A simple method of sentiment analysis classifies reviews of user's into positive, negative or neutral. Whenever review of user expresses a positive opinion, it is denoted by positive label. Similarly, if review expresses a negative opinion than it comes under negative label.

But, Sentiment analysis has many challenges. There can be structured, unstructured and semi-structured data coming from multiple sources into web. The sentiment analysis has to identify the expressions of opinion and mood of writers from variety of data. The complexity of data become major issue when users involve images, emoji, symbols etc. Sometimes, a word which considered as positive in one situation can be considered negative in another situation. Many reviews may contain both positive and negative comments for a product or service. For humans, it will be easy to break up a sentence and analyse its meaning. But, to implement it programmatically is also a difficult job. Since people have different mindset everyone expresses opinions in a different way. Hence, to design intelligent sentiment analysis system is a real challenging task.

Sentiment Analysis has been used in range of different areas from including marketing, politics, healthcare, media and psychology. Sentiment analysis can be used to analyse trends, identify ideological bias, target advertising and gauge reactions [1].

Twitter is a social networking site that allows users to post short messages that must be under 140 characters. This makes twitter a microblogging website. These messages are known as "tweets". The appeal of twitter for sentiment analysis lies in the popularity of the site. According to *Internet Live Statistics*[2] every second, on average, around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year. This number of tweets provides a huge dataset to gauge public opinion on a number of different issues. Another benefit of twitter is that by using the twitter API(Application Interface Program), live tweets can be pulled from twitter. This provides the ability to analyse the public opinion in real time.

Sentiment analysis is a growing field and is applied on twitter data for providing benefits in various areas. Section II provides literature review which helps to identify some good research work in the field of sentiment analysis.

## II. RELATED WORK

Twitter datasets need a different methodological approach in order to overcome the size and complexity issues. Kimmons and Veletsianos [3] focus on the tweeting activity of the American Educational Research Association annual conferences between 2014 and 2015, inferring that participants with different roles have different behaviour. The same research applied sentiment analysis to the conference's hashtags, concluding that academics generally exhibit positive sentiments. Desai et al. [4] also tried to detect sentimental polarity at a medical conference's tweets and found that, in

general, informative tweets had more negative sentiment scores than uninformative tweets. They also found that these were more negative during conferences than in the leading up period.

Jaspreet Singh et. al.[5] worked on four state-of-the-art machine learning classifiers viz. Naïve Bayes, J48, BFTree and OneR for optimization of sentiment analysis. The Naïve Bayes found to be quite fast in learning whereas OneR seems more promising in generating the accuracy of 91.3% in precision, 97% in F-measure and 92.34% in correctly classified instances.

Vishal et al. [6] reviewed current procedures for opinion mining such as machine learning and vocabulary-based methodologies. Utilizing different machine learning algorithms like NB, Max Entropy, and SVM, Vishal et al. [6] additionally described general difficulties and utilizations of Twitter sentiment analysis.

Mihai Dascălu, et.al, proposed an automatic approach with respect to NLP and used distributing figuring to improve the runtime performance [7]. Further, with respect to the multilayered engineering an exceptional grading component is provided. A replicated worker design is sent to improve the speed of this process. In this approach, along with increment in performance level, there are two important aspects to be considered which are, load balancing and fault tolerance. The corpus of chats can be assessed in a timely manner resulting is providing a quick access to the participants' feedbacks as per this demonstration of framework. By investigating huge corpuses in very less time, a solution is provided through which the general performance is improved. This is done by utilizing the subtle elements on a distributed form of instrument. Under different conditions and loads, the performance of proposed work is better as compared to existing approaches.

Go and L.Huang [8] proposed an answer for conclusion examination for Twitter information by utilizing far off supervision, in which their preparation information comprised of tweets with emojis which filled in as uproarious names. Go et al [8] introduced a method to classify the sentiment of tweets. The idea behind it was to aggregate feedback automatically. The sentiment problem was treated as a binary classification, in which tweets were classified into positive and negative. Training data containing tweets with emoticons were collected based on supervision approach that was proposed by Read [9]. To achieve this, Go et al [8] utilized the Twitter API to extract tweets that included emoticons. These were used to identify tweets as either negative or positive. Retweeted posts and repeated tweets were removed. In addition, tweets containing positive and negative emotions were filtered out. Various classifiers such as the NB, MaxEnt, and SVM were employed to classify tweets. Different features were extracted such as unigrams, bigrams, unigrams with bigrams, and unigrams with POS. The best results were obtained by the MaxEnt classifier in conjunction with unigram and bigram features, which achieved an accuracy of 83% compared to the NB with a classification accuracy of 82.7%.

Malhar and Ram [10] proposed the supervised method to categorize Twitter data. The results of this experiment demonstrated that the SVM performed better than other classifiers and, using a hybrid feature selection, achieved an accuracy of 88%. The experiment attempted to combine principal component analysis (PCA) alongside the SVM classifier to reduce feature dimensionality. Furthermore, unigram, bigram, hybrid (unigram and bigram) feature-extraction methods were used. Malhar and Ram [10] showed that integrating PCA with the SVM with a hybrid feature selection could help in reducing feature dimensions and the results obtained a classification accuracy of 92%.

Anton and Andrey [11] developed a model to extract sentiment polarity from Twitter data. The features extracted were words containing n-grams and emoticons. The experiment carried out demonstrated that the SVM performed better than the Naïve Bayes. The best overall performing method was the SVM in combination with unigram feature extraction, achieving a precision accuracy of 81% and a recall accuracy of 74%.

Jim Samuel et. al. [12] experiment over Covid-19 tweets proved that Naïve Bayes is better for small to medium size tweets and can be used for classifying short Coronavirus Tweets sentiments with an accuracy of 91%, as compared to logistic regression with an accuracy of 74%. For longer Tweets, Naïve Bayes provided an accuracy of 57% and logistic regression provided an accuracy of 52%

### III. CONCLUSION

The dichotomy of sentiment is generally decided by the mindset of an author of text whether he is positively or negatively oriented towards his saying [12] to classify words into their respective categories. The present study

shows methods used in the field of sentiment analysis. Twitter provides valuable informational and public sentiment insights which can be used to develop much needed motivational solutions and strategies in different problem areas.

#### REFERENCES

- [1] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web," *Management science*, vol. 53, no. 9, pp. 1375-1388, 2007.
- [2] <https://www.internetlivestats.com/twitter-statistics/>
- [3] Kimmons, R., "Expansive openness in teacher practice", *Teachers College Record*, 118(9), 2016.
- [4] Desai, T., Shariff, A., Shariff, A., Kats, M., Fang, X., Christiano, C., & Ferris, M. (2012). Tweeting the meeting: An in-depth analysis of Twitter activity at Kidney Week 2011. *PLoS One*, 7(7), e40253, 1-9. <https://doi.org/10.1371/journal.pone.0040253>.
- [5] Jaspreet Singh, Gurvinder Singh and Rajinder Singh, "Optimization of sentiment analysis using machine learning classifiers", *Hum. Cent. Comput. Inf. Sci.* (2017) 7:32.
- [6] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," *arXiv preprint arXiv:1601.06971*, 2016.
- [7] Mihai Dascălu, Ciprian Dobre, Ștefan Trăușan-Matu, Valentin Cristea. Beyond Traditional NLP: A Distributed Solution for Optimizing Chat Processing. 2011 10th International Symposium on Parallel and Distributed Computing.
- [8] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report*, Stanford, vol. 1, no. 2009, p. 12, 2009.
- [9] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL student research workshop, 2005: Association for Computational Linguistics*, pp. 43-48.
- [10] M. Anjaria and R. M. R. Guddeti, "Influence factor based opinion mining of Twitter data using supervised learning," in *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*, 2014, pp. 1-8.
- [11] A. Barhan and A. Shakhomirov, "Methods for Sentiment Analysis of twitter messages," in *12th Conference of FRUCT Association*, 2012.
- [12] Jim Samuel, G. G. Md. Nawaz Ali, Md. Mokhlesur Rahman, Ek Esawi, and Yana Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", doi:10.20944/preprints202005.0015.v1