

An Exploration of Sentiment Analysis using Twitter Dataset

Dr. Jyothi Mandala¹, Pragada Akhila², Uriti Archana³

¹ Assistant Professor, Department of CSE, School of Engineering & Technology, CHRIST (Deemed to University), Kengeri Campus, Bangalore, India.

² Assistant Professor, Department of CSE, Vignan Institute Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

³ Assistant Professor, Department of IT, GMR Institute of Technology (A), Rajam, Andhra Pradesh, India.

ABSTRACT

With the development and growth of web technology, there is a large volume of data present for internet users on the web, and data is also generated. The Internet is now an online learning platform where you can exchange your opinions and share your valuable ideas. Social media and websites like Facebook and twitter are becoming more popular, allowing people to share their views and ideas on different subjects, talk with different people with different cultures or sharing messages around the world. Mainly when it comes to Twitter the data sentiment analysis a lot of work has been present. A tweet, is small and an expression of necessary ways and is typical. We are presenting our review paper here and in this, we concentrated on Twitter information analysis of sentiment. Here the text mining and NLP are seen as the sentimental analysis. Research of different aspects on twitter data sentiment can be carried out. This article shows how sentiment analysis styles and methods are used for extracting emotions from tweets. We've got a differentiative study of multiple methods and ways of approaching for analysis of sentiment with data of twitter in this survey paper.

Keywords

Sentiment Analysis, Opinion Mining, Social Media, Twitter Data.

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

Twitter, Facebook, and YouTube and other social media platforms have become so popular over the past few years. The study of emotions in this field is known as opinion mining and is covered by data mining and computer linguistics. The detection of mood is its primary objective. Behaviour and opinions of the person from the documents. The increased use of sites for social media has contributed to the use of publicly available data for sentiment analyses through sentiment analysis methods in various fields such as economics, the finance, and sociology. Mostly the unstructured data is present on the social network .80% of the world's results are unstructured only. It increases the difficulty of interpreting and extracting reliable conclusions from such information. Sentient mining of data or opinion is an effective technique to detect people's views on data of social platforms.

Opinions which are taken from others may be applicable when a decision is made. When these choices require valuable resources, people consider the past experiences of their partners. It is now became easy to post our ideas easily with different people. Even though the polarity detection is the only one the study of sentient

focused on positives, negatives, or neutral. Twitter is a micro-blogging platform with a wide variety of short-term applications for social networking and advertising. Political parties, for instance, might like to ask whether or not people support their curricula. In this scenario, it is necessary to gather opinions on social platforms and to make conclusions on what people admire, which is the most important aspect.

SENTIMENT ANALYSIS

The opinion mining or sentiment analysis is an area of research, which interprets the opinions of individuals against any specific subject, events, etc. in text mining. It is a problematic area. There are other different names and functions, like sentiment analysis, opinion mining, subjectivity analysis, mining evaluation, etc.

Analysis levels:

Sentiment analysis is usually divided into three grades:

1. Document Level Analysis:

Each rate classifies the whole report whether it gives a positive or a negative feeling. The paper considers a single topic. Texts that include

comparative training cannot, be considered at the text level.

2. Analysis of sentence level:

In this level the task is to decide whether each sentence has a positive, negative or a neutral opinion. An unbiased sentence which doesn't send an opinion is a neutral opinion. The analysis of sentence-level is based on the classification of subjectivity. It consists factual information from subjective phrases and sentiments.

3. Entity/Aspect Level Analysis:

Both the text and the sentencing stage studies doesn't like or dislike people. The level of entity/aspect basically offers insight. This level was previously called the level of the function. The central task of the organizational level is to define structures. The aspect level places direct attention on opinions or feelings. It is based on the idea that a view exists according to its behaviour and its intent.

RELATED WORK

There are different text mining methods in order to mine the data.

PSingh et al, Used the sentiment analysis approach to Examine the government policy on demonization from the stock point of view, they collected Twitter data using specific hashtags (# demonetization) and also Geo-location analysis (tweets received statewise). The feeling analysis API from the cloud meaning classified different states into six categories; as pleased, sad, very sad, very happy, neutral, and no results.

Justin et al, Solved the question of the categorization of sentiment polarity. It is one of the basic sentiment analysis problems. Data was collected from the Amazon.com and were used to analyse the online product reviews. In this paper, research is carried out by the categorization of both sentences and the reviews. Scikit-learn code is used for this analysis. It is a software package for an open-source education in Python. Classification techniques used for categorization are Naïve Bayesian, Random Forest, and SVM.

Yadav et al, They contributed to the customer review classification sentiment analysis. Labelled Twitter data was used in this task. In this paper,

they used three techniques under supervision: naïve-Bayes, Max-entropy, and SVM, accompanied by semanticized analyzes. Using all the three methods for estimating similarities, they have trained and classified Python and NLTK: naïve-Bayes, Max-entropy, and SVM. Better outcomes were given by Naïve-Bayes approach than the Max-entropy and SVM with a Unigram model. The accuracy is improved when the semantic analysis Word-Net applied after the above process.

M S Nee et al. The Machine Learning Process was used to analyze the Twitter data on Digital Products. A new feature vector exists for classifying the tweets and providing opinion of the people about electronic products. It creates a Feature-Vector from eight features. The eight features used are the particular keyword, negation presence, and several positive keywords, emoticons, neutral keywords, negative hashtags, and positive hashtags. Using built-in functions to implement Naïve-Bayes and SVM classifiers. The classifier Max-Entropy is applied with the Maximum-Entropy code. All classifiers are equally efficient.

Jivane et al. suggested a more accurate model of twitter sentiment analysis data on reviews of Hollywood and Bollywood films. We classify these tweets accurately by using classifiers and feature vectors such as SVM and Naïve-Bayes based on the feeling of every tweet. Naïve-Bayes is significantly accurate than SVM, but with significantly less precision and alert. But based on precision SVM is superior to Naive-Bayes. The Among the selected classifiers Feature Vector provides a better feel analysis. The accuracy in Classification will increase with the increase in the training data.

TWITTER

The aim of the study of twitter sentiment is to correctly identify the tweets in different sentiment groups. Various techniques have been developed in this area of research which provides various methods of training a model and tests to check its effectiveness. Sentiment analysis on twitter tweets is quite challenging. Here we describe a few explanations.

Small volume of tweets:

Condensed statements are produced with only 280 characters in hand, resulting in scarce features.

Use of jargon:

These words are not precisely the same as the English words, we regularly use but because of the historical use of slangs, they make the dated approach.

Features of Twitter:

It allows the usage of users, hashtags, and URLs. Which require various approaches and treatments to be compared to the other words.

Variety of clients:

Users express their views in various forms; some use different languages, while others use repeated words or symbols to express their emotions.

SENTIMENT ANALYSIS ON TWITTER DATA

Figure.2 demonstrates the process of sentiment analysis. The framework includes the four main modules:

1. Data collection module,
2. Data processing module,
3. Grading module,
4. Performance analysis.

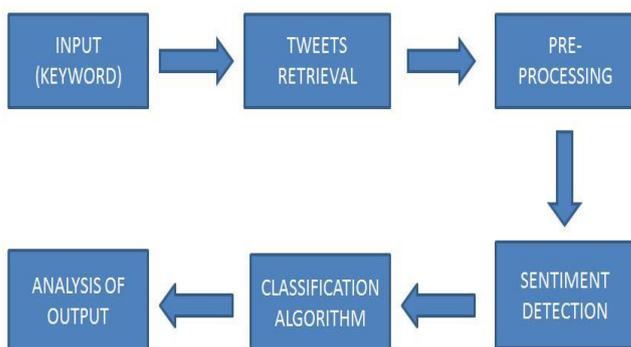


Fig 1: Work Flow of Sentiment Analysis

a. Input (Keyword):

We choose a topic and collect the tweets with the keyword and examine the feelings on these tweets.

b. Tweets Retrieval:

Tweets can be differentiated into three types Structured, semi-structured, and unstructured. We can compile tweets using different programming languages such as R or python and research on sentiment analysis.

c. Pre-Processing:

Sorting the data to remove incomplete noisy and inconsistent data is called pre-processing data. Pre-processing tasks include the following functions:

- The Elimination of tweets
- Deleting the URLs, special characters, lines, numbers, etc.
- Removal of the stopwords
- Stemming

d. Sentiment Detection:

In applications of sentiment analysis and opinion mining, feeling word recognition is essential to work. Such as tweeter mining, opinion holders finding, and the tweet classification. Through the process of Sentiment Analysis, the fundamental task is to identify the polarity of the particular tweet, The polarity i.e., Negative and positive. Negative Polarity recognition is carried out by using different lexicons, for example, Bing Lui lexicon feeling, SentiWordNet, etc. They help in calculating the feeling strength, feeling score, etc.

e. Classification Algorithm:

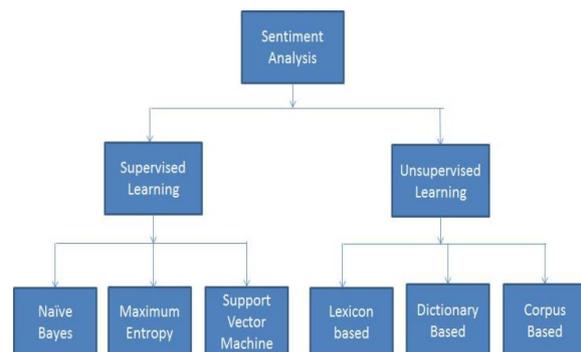


Fig 2: Algorithms of Sentiment Analysis

There are two main methods of psychological evaluation, i.e., Supervised approach to education and the non-supervised approach to learning. Twitter data is categorized into different forms of supervised machine learning methods such as Naïve-Bayes, SVM, and Maximum Entropy, etc.

Classifier efficiency is hugely dependent on which dataset is used for the classification methods. For supervised machine learning methods, the Training dataset for the classification system is used in helping to classify test data.

f. Analysis of Output:

The basic idea of sentiment analysis is to change unstructured data into usable or structured data, which is meaningful. After the review is

completed, the results shown on graphs in the form of diagrams, bar diagrams, and line graphs.

COMPARATIVE STUDY OF TECHNIQUES OF SENTIMENT ANALYSIS HAVING TWITTER DATASET

The table below displays the work of various authors on Twitter data set sentiment analysis.

Author & Year	Dataset	Techniques	Accuracy
Geetika Gautam (2014)	Customer Review Twitter Dataset	Naive Bayes Maximum Entropy SVM Semantic Analysis (WordNet)	88.2% 83.8% 85.5% 89.9%
Nee M. S. (2013)	Twitter posts about electronic products	Navive Bayes SVM Maximum Entropy Essembled	89.5% 90% 90% 90%
Seyed-Ali Bahrainian (2013)	Twitter data on Smartphones	Unigram feature, SVM, NB, MaxEnt Hybrid Approach	89.78%
Dhiraj Gurkhe (2014)	Twitter Data	Unigram Bigram Uni+Bigram	81.2% 15% 67.5%
Apoorv Agarwal (2011)	11,875 manually annotated Tweets	Unigram Senti-features Kernel Unigram + Senti-features Kernel + Senti-features	71.35 71.27 73.93 75.39 74.61

Table1: Summary of Twitter research articles

CONCLUSIONS

Twitter data is analysed in various angles to mine the opinion or feeling. This document defines the concept of sentiment analysis and opinion mining for different levels of sentiment analysis. This

paper discussed various methods of sentiment analysis and its methodologies.

It is necessary to know the Twitter, its structure, and its significance while doing a Twitter feeling study. This article gives a brief insight into tweets. The essential information for Twitter sentiment analysis is well discussed in this review paper.

A literary study shows that accuracy is improved when the WordNetsynanthropic analysis is followed by machine-learning techniques such as SVM, Naïve-Bayes, and Maximum Entropy. The reliability of the hybrid approach can also be improved by up to 4-5%.

REFERENCES:

- [1] Surnar, Avinash, and Sunil Sonawane. "Review for Twitter Sentiment Analysis Using Various Methods."
- [2] Eliacik, AlpaslanBurak, and ErdoğanErdoğan. "User-weighted sentiment analysis for financial community on Twitter." *Innovations in Information Technology (IIT)*, 201511th International Conference on.IEEE, 2015.
- [3] Ahmed, Khaled, Neamat El Tazi, and Ahmad HanyHossny.
- [4] "Sentiment Analysis over Social Networks: An Overview." *Systems, Man, and Cybernetics (SMC)*, 2015 IEEEInternational Conference on.IEEE, 2015.
- [5] Ko, Youngjoong, and JungyunSeo. "Automatic text categorization by unsupervised learning." *Proceedings of the18th conference on Computational linguistics-Volume 1*.Association for Computational Linguistics, 2000.
- [6] Kharche, S. R., and LokeshBijole. "Review on Sentiment Analysis of Twitter Data." *International Journal of Computer Science and Applications* 8 (2015).
- [7] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
- [8] Singh, Prabhsimran, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government." *ICT Express* (2017).
- [9] Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data." *Journal of Big Data* 2.1 (2015): 5.
- [10] Gautam, Geetika, and DivakarYadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." *Contemporary computing (IC3)*, 2014 seventh international conference on. IEEE, 2014.
- [11] Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." *Computing, Communications and Networking Technologies (ICCCNT)*, 2013 Fourth International Conference on.IEEE, 2013.
- [12] Amolik, Akshay, et al. "Twitter sentiment analysis of movie reviews using machine learning techniques." *International Journal of Engineering and Technology* 7.6 (2016): 1-7.
- [13] Gupta, Bhumika, et al. "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python." *International Journal of Computer Applications* 165.9 (2017).
- [14] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment Analysis of Events from Twitter Using Open Source Tool." (2016).
- [15] Bahrainian, Seyed-Ali, and Andreas Dengel. "Sentiment analysis and summarization of twitter data." *Computational Science and Engineering (CSE)*, 2013 IEEE 16th International Conference on. IEEE, 2013.
- [16] Gurkhe, Dhiraj, and Rishit Bhatia. "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification." (2014).