International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, March-April-2021, ISSN (Online): 2395-566X

Review of Twitter Data Sentiment Analysis

M. Tech. Scholar Nikita Akhand, Asst. Prof. Ms. Shivani Gupta
Department of CSE,
Patel Group of institutions, Indore
ernikitaakhand2610@gmail.com, Shivanigupta@patelcollege.com

***Abstract*- Everyday a huge amount of data has been produced over various social sites. A huge number of users share and tweet regular updates on twitter. Tweet is a short way of expressing thoughts on any topic. So, the sentiment analysis of this data is must to keep track of the tweets. Sentiment analysis is the way to carry out text mining. To analyze this data produced over the twitter there are many methods that are used. Various experiments have been done to perform sentiment analysis of this data produced over twitter. There are different levels on which the sentiment analysis can be performed. So, in this paper we are providing a survey on the sentiment analysis of twitter data which uses the supervised and unsupervised learning algorithms.**

***Keywords*- Sentiment Analysis, Social Media, Twitter data, Supervised learning.**

I. INTRODUCTION

Social networking sites like facebook twitter etc are becoming so popular now days. The area of sentiment analysis is also known as Opinion mining. Opinion mining comes under the vision of data mining and computational linguistics. The main purpose of applying opinion mining is to detect the behavior, mood and attitude of the user. As the social sites took place in our life, sentiment analysis is used to study the reactions of users in different areas such as politics, economy, finance and sociology.

The data found over the social networks is unstructured data. This data is almost 80% of the unstructured data that is present all over the world. Because the data is unstructured it is difficult to analyze and gain useful information form the available data. Sentiment analysis is the method to read the opinions of the user by textual mining placed over the network [1].

Fig 1. Social media for sentiment analysis [1]

Twitter is a micro blogging site which contains large number of short lengths utilizes for marketing, social networking [1]. When it comes to get opinions of peoples over any topic then this data is useful for making decisions. For example, various political parties may use this data to make predictions about their results based on people's tweets. Opinions of peoples are useful when it comes to make any decision.

Now a days to get perspective of peoples over different issues and to get the conclusions form their views have become an important way to gain information over any topic. Here in this paper we are going to provide a survey over the sentiment analysis of twitter data.

II. SENTIMENT ANALYSIS

Sentiment analysis is the area in which the data is analyzed to get the opinions of the users over any topic or any individual event etc. it is also known as opinion mining [4]. This mining uses text data to make the decisions and carry out the results. Various names are given to this mining i.e. opinion mining, sentiment mining, affect analysis, subjectivity analysis, review mining etc.

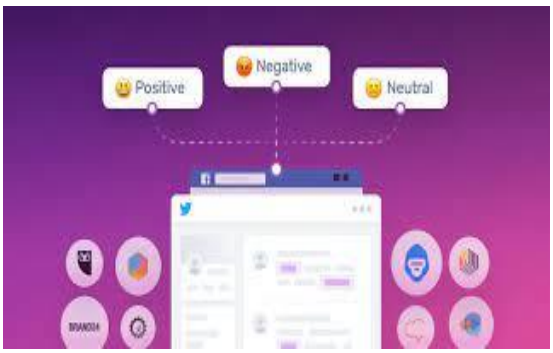

Fig 2. Sentiment Analysis. [1]

Different levels of analysis:
The sentiment analysis can be carried into three different levels that are:

© 2021 IJSRET

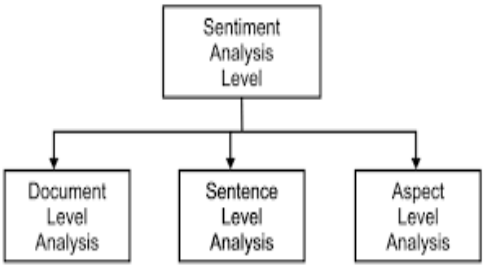

Fig 3. Levels of Sentiment Analysis.

1. Document Level Analysis:
Document level analysis performs the analysis over the whole document and obtains the result that the document gives a positive sentiment or negative sentiment. In this level the whole document is considered on a single topic. So the document which contains comparative learning can not be considered for document level analysis.

2. Sentence Level Analysis:
This level of analysis gives the opinion, sentence by sentence. Here every sentence is analyzed to give the sentiment i.e. positive, negative or neutral. When the sentence does not provide any opinion then it is termed as neutral sentiment. This analysis provides subjective level sentiments that the sentence is good or bad.

3. Entity/Aspect Level Analysis:
Document and sentence level analysis does not provide any information about the likes and dislikes of the peoples. This level gives direct information about the attitude about the opinion. This level is also called as feature level analysis.

III. RELATED WORK

Authors proposed a model to work on real time data produced over twitter. The proposed model varies from the earlier work done In this area as it use both the supervised and unsupervised machine learning algorithms together. To show the popularity of the two shops Mcdonald and KFC the data was collected for this two. A result shows that McDonald is more popular as compared to KFC. [2]

Twitter is used mainly as it has got a lot of attention. Direct responses were used from peoples as this makes the data more real. In this paper, hashtags are used to find out the interest of peoples and to sort the topics. The data that is cleaned was applied on following machine learning techniques, i.e. naïve bayes, random forest and tailored random forest. Results shown that random forest have more sentiment accuracy as compared to naïve bayes. [3]
In this paper, by using R the collected tweets are pre-processed, analyzed and visualized and for this author presented an open-source approach. R programming tool is used to for sentiment analysis. This experiment gives eight different classification feelings and the two sentiments. They used approx 3000 tweets to do this classification. [4]

In this paper the authors proposed a sentiment analysis based on visual framework that is used for the analysis of Huawei technologies when the trade war between US-China happened. Sentiment analysis and the geographic visualizations were used in the proposed model. The obtained results were usefull for different economists and politicians to guess the public opinion. [5]

To do sentiment analysis over twitter data the authors compared different deep learning techniques. Evaluation and comparison of CNN and RNN (Long Term Short Memory) networks were done. The results show the limitations of CNN and Long-Term Short Memory as their low performance in the field. [6]

Sentiment diffusion patterns and the textual information were used to get results and to improve performance of the sentiment analysis performed over twitter data. The use of both the textual information and diffusion patterns were made as they are related at some context. The algorithm proposed by authors named as SentiDiff to make sentiment prediction over twitter. The proposed algorithm achieves PR-AUC increments between 5.09% and 8.38%. [7]

Authors presented a system for real ttime sentiment analysis on twitter that works online as it provides real time system. The graphical representation of the polarity of tweets can be carried out by giving input query. Simple voter and naïve bayes algorithms are used for the classification of tweets. The proposed 3 way classifier was analyzed by using the simple voter and naïve bayes algorithms. [8]

In this paper authors used reviews about the mobile brands; they collected more than 5000 reviews from the twitter platform. After that the applied lexicon approaches to perform sentiment analysis. From this analysis they declared that among all brands the Motorola, Samsung and oppo are more popular inn market. [9]

The authors used the tool Rapid Minerto get the data from twitter, after this they used AYLIEN extension, preparing the dataset and analyzing the sentiments. The sentiments for the president was analysed which are classified on three basis i.e. positive, negative and neutral. [10]

In this paper authors presented a technique to classify the sentiments on twitter by using machine learning domain, which uses various textual patterns for example n-grams of twitter data. To analyze the proposed model the performed experiments of 4 sets of n-gram features and 3 weighting techniques. The results examined that unigram features performs very well as compared to others. [11]

© 2021 IJSRET

To perform tweet segmentation, sentiment analysis and event detection authors proposed SEG Analysis framework. This method works on real time and gets online tweets and founds meanings of the candidates. They combined the two mechanisms that are Naïve Bayes and online clustering, to know the unknown events over period of time. [12]

Classification techniques were used to perform sentiment analysis. Firstly, they applied the SVM classifier and after proceeding with this the replaced it with KNN classifier, which increased the accuracy. The overall work was implemented in anaconda and results show that proposed method gives higher accuracy than SVM. [13]

Authors performed analysis of the reactions of peoples on the topic of GST, this analysis was done on the data of 100 days over a platform like twitter. Major Politian areas of country were considered to make analysis and they used R to perform this. The results stated that the nations are having equal response for this in terms of negative and positive sentiments. [14]

In this paper authors analyzed the sentiments of users for trend, brand, product or person. The proposed application was able to retrieve the word cloud for the frequently occurring words in tweets. The visualization was done by plotting histogram and by pie chart for the classifications for the calculated sentiments scores. [15]

In this paper the authors conducted a survey on the sentiment analysis and they gave a detailed study of the extraction of sentiments from tweets. They made a comparison of various methods and approaches that are used to perform sentiment analysis on twitter data. [16]

IV. TWITTER

When we perform the sentiment analysis, we need to classify the data and the results appropriately. The results obtained from the sentiment analysis are sorted in various types i.e. positive, negative and neutral. To perform these various types of techniques are involved and various experiments have been performed till now and the work is still going on. When it comes to perform this analysis on twitter data many challenges have been faced.

Some of them are:

1. Size of Tweet:
The tweet carries only 280 characters to present the thought. So, by having this small number of characters, it becomes typical to evaluate the data.
2. **Slang in Tweet:** When the slang is used in the tweet it becomes more typical to derive the meaning of the words because these words are not same as the words of English. Slangs make the task more challenging.

3. Special Symbols: Twitter provide the use of special symbols like hashtag, URL etc, which needs special techniques to evaluate the meaning of tweets.

4. Variety in Tweets: The tweets done by users belong to different locations, languages, mood and different time. Somewhere tweets carry symbols, pictures to express their feelings.

V. SENTIMENT ANALYSIS OF TWEETS

Sentiment analysis comes with the text mining for which different methods are used. On twitter more than 400 million tweets are done per day. So, to perform sentiment analysis on twitter data becomes more important. It also provides real time data to evaluate the sentiment.

The steps performed to evaluate the twitter data using sentiment analysis are:

1. **Data Gathering:** It can be done in two ways i.e. current tweets and historical tweets.

2. **Data Preparing:** In this step we perform some task such as removing irrelevant information, special characters and blank spaces.

3. **Creation of Sentiment Analysis Model:** In this step we create a model to perform analysis on the obtained data. Evaluation of this data is done with this created model.

4. **Visualization of the Results:** In this step we see the results in a very good visual. This step provides the obtained results in an impactful manner.

VI. CONCLUSION

Opinion mining or sentiment analysis comes with a great scope for examining the people's opinion over any topic. It provides a way to increase the business or to predict the results based on the opinions.

Twitter gives a big platform for users to provide their views on any topic. So, this data produced over twitter in the form of tweets is used for sentiment analysis. It provides the sentiment of users in positive, negative or neutral basis. The various supervised learning algorithms studied in this survey states that the results obtained from this algorithm varies between 68 to 84%. The supervised learning approaches studied are N-gram, Naïve Bayes, SVM, and Random Forest. Also, while performing the sentiment analysis there are many challenges that are seen and discussed in this paper. So, we tried to give a survey over the sentiment analysis of twitter data.

© 2021 IJSRET

REFERENCES

[1] El Rahman, S. A., AlOtaibi, F. A., & AlShehri, W. A. (2019). Sentiment Analysis of Twitter Data. 2019 International Conference on Computer and Information Sciences (ICCIS).
[2] Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi and Wejdan Abdullah AlShehri, Sentiment Analysis of Twitter Data, 978-1-5386-8125-1/19/$31.00 ©2019 IEEE.
[3] Rekha V, Raksha R, Pradnya Patil, Swaras N and Rajat GL, Sentiment Analysis on Indian Government Schemes Using Twitter data, 978-1-5386-9319-3/19/$31.00 ©2019 IEEE.
[4] Sonia Saini, Ritu Punhani, Ruchika Bathla and Vinod Kumar Shukla, 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University, Sentiment Analysis on Twitter Data using R.
[5] Nann Hwan Khun and Hninn Aye Thant, Visualization of Twitter Sentiment during the Period of US Banned Huawei.
[6] Sani Kamış and Dionysis Goularas, 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data.
[7] Lei Wang, Jianwei Niu, and Shui Yu, SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis. JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2018.
[8] Alaa S. Al Shammari Real-time Twitter Sentiment Analysis using 3-way classifier, 978-1-5386-4110-1/18/$31.00 ©2018 IEEE.
[9] Ms. Minu Choudhary and Mr. Prashant Kumar Choudhary, International Conference on Smart Systems and Inventive Technology (ICSSIT 2018), Sentiment Analysis of Text Reviewing Algorithm using Data Mining, 978-1-5386-5873-4/18/$31.00 ©2018 IEEE.
[10] Merima Čišija, Emir Žunić and Dženana Đonko, Symposium on neural networks and applications 2018, Collection and Sentiment Analysis of Twitter Data on the Political Atmosphere, 978-1-5386-6974-7/18/$31.00 ©2018 IEEE.
[11] Sheeba Naz, Aditi Sharan and Nidhi Malik, 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), sentiment classification on twitter data using support vector machine, 978-1-5386-7325-6/18/$31.00 ©2018 IEEE.
[12] Mamta Patil and H.K. Chavan, Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018), Event Based Sentiment Analysis of Twitter Data.
[13] Reeta Rani and Sawal Tondon, 2018 4th International Conference on Computing Science, Chat Summarization and Sentiment Analysis Techniques in Data Mining, 978-1-5386-8025-4/18/$31.00 ©2018 IEEE.
[14] Shubham Roy, Samarth Sehgal and Shantanu Agrawal, International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018) ISBN: 978-1-5386-4119-4/18/$ 31.00 ©2018 IEEE 32 An Approach to Sentiment Analysis of Twitter Data on the Goods and Services Tax.
[15] Prakruthi V, Sindhu D and Dr S Anupama Kumar, 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018 ISBN: 978-1-5386-6078-2 © 2018 IEEE 29 Real Time Sentiment Analysis of Twitter Post.
[16] Rasika Wagh and Payal Punde, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) Survey on Sentiment Analysis using Twitter Dataset.

© 2021 IJSRET