International Journal of Computer Science Engineering & Technology
APPROVED BY
NATIONAL SCIENCE LIBRARY (NSL)
NATIONAL INSTITUTE OF SCIENCE COMMUNICATION AND INFORMATION RESOURCES (NISCAIR)
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH (CSIR)- NEW DELHI INDIA .
ISSN :2455-9091

A SURVEY ON SENTIMENT ANALYSIS OF TWITTER DATA USING DIFFERENT TECHNIQUES

[1] J. Uma, [2] Dr. K. Prabha,
[1] Research scholar, [2] Assistant Professor,
[1, 2] Department of computer science,
[1, 2] Periyar University PG Extension centre,
[1, 2] Dharmapuri.

ABSTRACT - Conceptual - Recent years have seen the fast development of social media stages (e.g., Facebook, Twitter, Google+, and a few sites) wherein clients can distribute contemplations and opinions on any subject. Sentiment analysis in Twitter is a field that has as of late pulled in research interest. Twitter is one of the most well known microblog stages on which clients can distribute their musings and opinions. Sentiment analysis in Twitter handles the issue of examining the tweets as far as the supposition they express. This review gives a diagram of the subject by examining and quickly depicting the algorithms that have been proposed for sentiment analysis in Twitter.

Keywords: [Facebook, Twitter, Opinions, Sentiment Analysis, Tweet]

1. INTRODUCTION

Sentiment Analysis

Sentiment analysis is the proportion of individuals' opinions fair and square of concurrence on a particular point, an item, or an assistance, or even elections. Two methodologies had been utilized to consider the sentiment analysis: natural language processing, and machine learning algorithms. To evaluate the clients' opinions in the past some paper-based reviews had been utilized, however it is hard to monitor and gather every one of clients' opinions. With the expanding wonders of social media it has become simpler and more open to creep every one of clients' feedbacks and investigate their sentiments as sure or negative.

Natural Language Processing Approach

Natural language processing (NLP) is the connection among PCs and human (natural) languages. To assess sentiment of clients on the web, especially on twitter, successful sentiment annotation ought to be utilized. Most examinations utilize the three normal sentiment names: positive, nonpartisan, and negative. New element had been utilized to viably explain sentiments of clients; "Blended Sentiment name", it exists in tweets that have two unique implications. For instance "I love iPhone, however I scorn iPad". "iPhone" element is commented on with positive sentiment name, and "iPad" element is explained with negative sentiment name, that implies the tweet has a blended sentiments.

Twitter

Twitter is our expect to order the tweets in different sentiment classes to perform sentiment analysis on twitter information. In this examination field, to prepare a model a few methodologies have created which is likewise utilized for testing to check its efficiency. It is trying to perform sentiment analysis on twitter information. There are a

few reasons characterized for this. Bound tweet Sizesize of the tweets is limited for example it is having just 280 characters which create serious proclamation. By using slang wordsthese slang words are bit unique in relation to English words. On the off chance that we utilized slang words in our sentence, at that point that slang words make a methodology antiquated. Twitter highlights twitter is tolerant to utilization of URL's, hashtag(#) and client reference. In the twitter highlight various cycles are utilized to contrast and other various words. Client variety there are different approaches to pass on their assessment, individuals' utilized distinctive language between the tweets while others utilized reprise words or images to pass on their emotions.

In this paper, we center around twitter for information analysis, where twitter is an internet organizing administration that empowers clients to send and peruse short 140-character messages called "tweets". Notwithstanding its exposure, twitter is available for unregistered clients to peruse and monitor most tweets, dissimilar to Facebook where clients can control the security of their profiles. Twitter is additionally an enormous social systems administration microblogging website. The enormous data gave by twitter, for example, tweet messages, client profile data, and the quantity of devotees/followings in the organization assume a huge part in information analysis, which consequently make most examinations research and look at different analysis methods to get a handle on the ongoing utilized advancements.

2. LITERATURE SURVEY

Framework for Sentiment Analysis in Twitter

Framework is a support structure or system that holds parts together, has something stretched over it or acts as the main structure.

1. Jenifer Jothi Mary A, Arockiam L (2017) proposed Jen-Ton Framework which is made to give precise analysis of the sentiment of the long reach casual communication data. It comprises of three novel strategies specifically, IMS, ASFuL and ASTA-CGA. The essential system is used to ascribe the missing sentiments of the corpus assembled from the online life in a legitimate method. This system utilizes three procedures specifically, Imputation of Missing Sentiment (IMS), Aspect based Sentiment Analysis using Fuzzy Logic (ASFuL) and Aspect based SenTiment Analysis using Clustering with Genetic Algorithm (ASTA-CGA) for updating the precision of the perspective based sentiment analysis in enormous data. A conclusive purpose of this paper is to propose procedures to update the precision of point based sentiment analysis and to develop a design system for improving the exactness of sentiment analysis in enormous data.

2. Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva, (2017) proposed another upgraded and all the all the more encouraging framework, as far as speed and productivity, for making meta-activities by executing Specific Action Rule disclosure reliant on Grabbing strategy (SARGS) count. For this, we play out a general analysis of metaactions making algorithmic execution in Apache Spark driven framework and customary Hadoop driven framework using the Twitter long reach casual communication data and assess the outcomes. We perform corpus based Sentimental Analysis of long reach interpersonal communication data, and test the absolute time taken by both the frameworks and their sub parts for the data planning. Results show faster computational time for Spark framework contrasted with Hadoop MapReduce for realizing the meta-movement age techniques. Data mining methods are used to dissect immense data sets, to perceive the major data designs and to reveal the disguised data. New algorithms have been proposed in the previous decade to find some exceptional activities reliant on the discovered examples as Action Rules. The Action Rules are unequivocal data designs eliminated from tremendous dataset which plans to change the present estimation of the versatile quality, viable, to an ideal worth.

3. Hossam Elzayady Khaled M. Badran Gouda I. Salama (2018) proposed an

effective sentiment forecast procedure, using the Apache Spark's Machine Learning library to execute diverse classification algorithms. The results show a significant upgrade in the precision of Naive Bayes and Logistic Regression with respect to increasing the volume of dataset, while the improvement is not strong in Decision Trees, also, examination's results reason that there is an inverse corresponding connection between running time and the quantity of machines in the Spark Cluster, So in case of adding additional nodes in the cluster, better capacity will be acquired. From the previous outcomes, our system can be described as powerful and scalable. Sentiment analysis can be described as a significant part of Natural Language Processing (NLP), its intend to distinguish the importance from a record so as to discover the extremity of the content [1-6]. For the sentiment analysis, we focus our consideration toward the Twitter, a microblogging social systems administration website, where users can speak with one another or share their opinions in short blogs.

4. Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, Erik Cambria (2017) proposed a phonetic-based framework for normalizing microtext to plain English and, henceforth, improve the classification precision of sentiment analysis. The proposed model consolidated two distinct methods for microtext normalization (specifically, vocabulary and phonetic-based) and Sentic API as an extremity classifier. 403 411 Ratcliff/Obershelp design coordinating calculation was seen as a well-suited assessment strategy. The results demonstrated that 85.31% of texts have a similarity record equivalent to or more noteworthy than 0.8, showing that this framework has the capacity to effectively deal with numerous types of normalization challenges. This paper proposes a novel framework to manage microtext normalization in Twitter in a human-inspired way, i.e., by shifting to the phonetic area to all the more likely disentangle microtext. Humans, truth be told, can understand abbreviations and uncanny spellings they have never seen because they naturally shift to the phonetic area when they read text.

Real Time Twitter Data Analysis

5. V. Prakruthi, D. Sindhu, Dr. S. Anupama Kumar (2018) presents a Systematic analysis of the tweets to give results was taken out from Twitter. Pack of words for positive and negative analysis were in kind of txt documents were taken to consider the tweets and request them into positive, negative and fair-minded tweets. This paper evaluates the individuals' sentiment about an individual, example, thing or brand. Twitter API is used to get to the tweets genuinely from twitter and assemble a sentiment classification for the tweets Sentiment analysis refers to the application for getting ready characteristic language, content analysis, computational semantics, and biometrics to purposely see, separate, assess, and learn brimming with feeling states and abstract data.

6. Alaa S. Al Shammari (2018) presents an online framework for constant twitter sentiment analysis and course of action. The proposed framework encourages clients to enter the request and get a graphical depiction of the tweets furthest point. Supposition mining and sentiment analysis is one of the most significant research subjects at our present time. It alludes to the investigation of individuals' feelings, sentiment or assessments that can be imparted in a composed book. Various organizations and associations endeavoring to discover the clients conclusion about their items, benefits, and so on to them for settling on better choices. What's more, Microblogging stages, for instance, Twitter, Facebook, weibo, etc can assist with separating and inspected customer's feelings and audits. The measure of these data is excessively enormous to possibly be destitute somewhere around common clients. Thusly, to avoid this, sentiment analysis methods could be used. The proposed framework assessed subject to the exactness of the recuperated tweets as demonstrated by the customer's data needs.

7. G. Kavitha, B. Saveen, Nomaan Imtiaz (2018) presents data spilling is a formative thought in enormous data where the size of data expands a ton from online social networking, slanting sites, and portable applications. These days, gushing data will all in all assemble data from live spilling to run analysis and produce reports for data desire. This method requires gifted professional for getting data from live stream using complex coding and inquiries. By using flume operator, the catchphrase record is put on Hadoop gathering to get appropriate data by means of flume station and thereafter sink the assembled data in Hadoop Distributed File System. The twitter data application makes a database in Hive and imports the assembled data to Hive table for visualization. It is basic for large business data stockrooms to device components to manage these data beneficially. Maker have displayed a structure for live gushing of twitter data for discovering general feelings in any space.

Phonetic based Analysis

8. Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, Erik Cambria (2017) proposed a phonetic-based framework for normalizing microtext to plain English and, henceforth, improve the classification accuracy of sentiment analysis. The proposed model joined two unique methods for microtext normalization (in particular, dictionary and phonetic-based) and Sentic API as an extremity classifier. 403 411 Ratcliff/Obershelp design coordinating calculation was seen as a well-suited assessment procedure. The results demonstrated that 85.31% of texts have a similarity record equivalent to or more noteworthy than 0.8, showing that this framework has the capacity to accurately deal with numerous types of normalization challenges. This paper proposes a novel framework to manage microtext normalization in Twitter in a human-inspired way, i.e., by shifting to the phonetic area to more readily decipher microtext. Humans, actually, can understand abbreviations and uncanny spellings they have never seen because they naturally shift to the phonetic area when they read text.

Fuzzy based

9. Ruchi Mehra, Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, Sunny Saxena, (2018) presents analysis for sentiment lead of Twitter data. The proposed work uses the guiltless Bayes and fluffy Classifier to orchestrate Tweets into constructive, adverse or neural direct of a specific person. It present preliminary assessment of our dataset and request results which demonstrated that consolidated proposed strategy is increasingly compelling as far as Accuracy, Precision and Recall. Sentiment analysis and assessment mining is the field that examine individuals' sentiments, conclusions, emotions from writings delivered by the clients. It is the dynamic investigation territories in like manner language preparing and is likewise extensively analyzed in web mining, data mining and online social networking assessment as sentiments are critical influencers of practices of human.

10. Padmaja K, Nagaratna P. Hegde (2019) presents that the system is produced for the classification of sentiment tweet classes. The blend of both ANFIS and GA accomplished superior execution, while identified with existing technique. The improvement approach (GA) with multiobjective classifier (ANFIS) moderately accomplished a classification accuracy of 92.78%. Furthermore, the proposed approach (ANFIS-GA) is prepared by using two fundamental phases; preprocessing and highlight age. This trial analysis was checked on an openly accessible database; twitter-sanders-apple2 that demonstrates the benefit of proposed strategy. Identified with existing approach (ANFIS-GA), the proposed philosophy accomplished a superior presentation considering accuracy, which showed 5.5-6% of improvement in twitter sentiment analysis. The ANFIS classifier was the fluffy based cosmology that was designed by realizing and the GA optimizes the fluffy principles in the ANFIS classifier.

Semi-Supervised Learning Algorithm

11. Metin Bilgin, Izzet Fatih Senturk (2017) presents system using a semi-supervised learning calculation on twitter messages. Two unique versions of the Doc2Vec calculation, DM and DBoW, which were as of late created and keep on taking a shot at the model, have been used. The success of the methods in the test phase was measured after the displaying. It is critical to pre-process, process and classify enormous measure of data appropriately. It is possible to classify the data acquired through social sharing sites by running various machine learning algorithms. For machine learning algorithms to have the option to classify data with high accuracy, the pre-processing phase must be accurately sorted out.

R programming

12. Sonia Saini, Ritu Punhani, Ruchika Bathla, Vinod Kumar Shukla (2019) presents an open source approach is presented which we have gathered tweets from Twitter API and afterward pre-processed, dissected and visualized these tweets using R. To examine sentiments of tweets we are using a statistical tool, R programming. This sentiment analysis is based on text data recovery from streamed web and afterward classifying individuals perspectives in eight distinct classifications of feeling (disgust, dread, outrage, expectation, sadness, trust, surprise) and two remarkable sentiments (positive and negative). The author describes the significance and applications of opinion mining and sentiment analysis in social networks and the basic concepts, challenges and comprehensive study in various sections . In this paper, an open source approach is presented which we have gathered tweets from Twitter API and afterward pre-processed, investigated and visualized these tweets using R. To break down sentiments of tweets we are using a statistical tool, R programming.

Author Name	Proposed Method	Merits	Demerits
Jenifer Jothi Mary A, Arockiam L (2017)	Proposed Jen-Ton Framework which is created to give precise analysis of the sentiment of the long range informal communication data.	Created architectural framework is improving the accuracy of sentiment analysis in enormous data.	(ABSA) space is the absence of a compelling architectural framework.
Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva (2017)	Propose another enhanced and all the more encouraging framework, as far as speed and productivity, for creating meta-activities by executing Specific Action Rule revelation dependent on Grabbing methodology (SARGS) calculation.	The attributes in a dataset are partitioned into adaptable attributes, whose worth is impermanent, and stable attributes, whose values is unchanging.	No robust and computerized technique for data sampling.
Hossam Elzayady Khaled M. Badran Gouda I. Salama (2018)	Proposed an efficient sentiment prediction technique, utilizing the Apache Spark's Machine Learning library	Apache Spark created to make processing and breaking down the data easier.	The social database the executives systems engines can't process unstructured or huge data any longer
Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, Erik Cambria (2017)	Proposed a phonetic-based framework.	To empower a similar fine-grained highlight based sentiment analysis. It is simple and instinctive to work. The processing time is genuinely short.	The analysis of multiword expressions that don't unequivocally pass on feeling.
V. Prakruthi ,D. Sindhu ,Dr. S. Anupama Kumar (2018)	Proposed an Systematic analysis of the tweets to give results was taken out from Twitter.	The application assembled was ready to produce word cloud for most incessant words found in the tweets recovered.	Computer programs have problems perceiving things like sarcasm and incongruity, negations, jokes, and exaggerations - the sorts of

		Visualization was among significant objectives, which was satisfied by plotting a histogram and creating pie outline for the classifications made based on sentiment score.	things a person would experience little difficulty recognizing.
Alaa S. Al Shammari, (2018)	Proposed an online framework for constant twitter sentiment analysis and arrangement.	The strategy that used is nonexclusive which can be used in different domains such as movies events.	Sometimes, the twitter API recovered less number of tweets than the number that requested. It is a very hard to break down these tweets based on misspellings, emoticon, and slang words where it should have a preprocessing step prior to managing it per the extremity discovery of positivity or negativity of the tweets.
G.Kavitha, B.Saveen, Nomaan Imtiaz, (2018)	Presents gushing data will in general gather data from live spilling to run analysis and produce reports for data expectation.	The user can see the rating yield for every single tweet brought from the watchword.	Supervised machine learning algorithms are proposed and their assessment shows that their accuracy is moderate and execution is low.
Ruchi Mehra, Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, Sunny Saxena, (2018)	The proposed work uses the innocent Bayes and fluffy Classifier to arrange Tweets into constructive, adverse or neural conduct of a specific individual.	It tends to be sensibly used as structures and applied for identification or classification tasks. It tends to be applied to huge sets of data to establish membership, in this case positivity, negativity and neutral.	In the event that all out factor has a classification in test data set, which was not observed in preparing data set, at that point model will assign a 0 (zero) likelihood and will be not able to make a prediction.
Padmaja K, Nagaratna P. Hegde (2019)	Presents that the system is developed for the classification of sentiment tweet classes. The combination of both ANFIS and GA achieved superior performance	This technique is used for recognizing the cluster heads from the twitter contents.	The length of the tweets is restricted that roughly 140 characters. Acknowledgment and distinguishing proof of nullify words from the tweets is a difficult task.
Metin Bilgin, Izzet Fatih Senturk, (2017)	presents system using a semi-supervised learning algorithm on twitter messages.	It is possible to classify the data acquired through social sharing sites by running various machine learning algorithms.	It is extremely hard to perform because it contains misspelled words, abbreviations and words and phrases from the day by day conversation jargon.
Sonia Saini, Ritu Punhani, Ruchika Bathla, Vinod Kumar Shukla(2019)	presents an open source approach	It will give away from of sentiment boundaries.	Without data assortment sentiment analysis is impossible and without data cleaning sentiment analysis is useless because that will be inaccurate.

CONCLUSION

The sheer amount and the various types of data on twitter and the public idea of tweets have permitted misusing twitter data in data analysis. Our point is to improve the analysis of twitter data for specific events to measure the impact and the behavior of users towards various events categories. A successive work will focus on studying the data and its attributes, and investigating modeling

techniques to recognize the recurrence distribution for every function. We have studied various approaches for sentiment analysis using various techniques. After that we organize the merits and demerits of the existing methods.

REFERENCES

[1]. Jenifer Jothi Mary A, Arockiam L (2017), "Jen-Ton: A Framework to enhance the Accuracy of Aspect level Sentiment Analysis in Big Data", ISBN: 978-1-5386-4031-9978-1-5386-4031-9/17/$31.00 ©2017 IEEE.

[2]. Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva,(2017), "Action Rules for Sentiment Analysis on Twitter Data using Spark", 2375-9259/17 $31.00 © 2017 IEEE DOI 10.1109/ICDMW.2017.14, 2017 IEEE International Conference on Data Mining Workshops

[3]. Hossam Elzayady Khaled M. Badran Gouda I. Salama (2018), "Sentiment Analysis on Twitter Data using Apache Spark Framework",
DOI: 10.1109/ICCES.2018.8639195,
Electronic ISBN: 978-1-5386-5111-7, IEEE.

[4]. Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, Erik Cambria (2017), "Phonetic-BasedMicrotextNormalization forTwitterSentimentAnalysis",
DOI: 10.1109/ICDMW.2017.59, Electronic ISBN: 978-1-5386-3800-2, IEEE.

[5]. V. Prakruthi ,D. Sindhu ,Dr. S. Anupama Kumar (2018)," Real Time Sentiment Analysis Of Twitter Posts",
DOI: 10.1109/CSITSS.2018.8768774,
Electronic ISBN: 978-1-5386-6078-2, IEEE.

[6]. Alaa S. Al Shammari , (2018), "Real-time Twitter Sentiment Analysis using 3-way classifier", 978-1-5386-4110-1/18/$31.00 ©2018 IEEE.

[7]. G.Kavitha, B.Saveen, Nomaan Imtiaz, (2018), "Discovering Public Opinions by Performing Sentimental Analysis on Real Time Twitter Data", 978-1-5386-0576-9/18/$31.00 ©2018 IEEE.

[8]. Ruchi Mehra , Mandeep Kaur Bedi , Gagandeep Singh , Raman Arora , Tannu Bala , Sunny Saxena,(2018), Sentimental Analysis Using Fuzzy and Naive Bayes", Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC), 978-1-5090-4890-8/17/$31.00 ©2017 IEEE.

[9]. Padmaja K , Nagaratna P. Hegde (2019), "Twitter sentiment analysis using adaptive neurofuzzy inference system with genetic algorithm", (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, IEEE

[10]. Metin Bilgin, Izzet Fatih Senturk, (2017), "Sentiment Analysis on Twitter data with SemiSupervised Doc2Vec",
DOI: 10.1109/UBMK.2017.8093492,
Electronic ISBN: 978-1-5386-0930-9, IEEE.

[11]. Sonia Saini, Ritu Punhani, Ruchika Bathla, Vinod Kumar Shukla(2019), "Sentiment Analysis on Twitter Data using R", 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University, 978-1-5386-8010-0/19/$31.00 ©2019 IEEE.