Survey on User Emotion Analysis using Twitter Data

Subramaniam.G

computer science dept,

IFET COLLEGE OF ENGINEERING

Cuddalore,india.

vinodhangv@gmail.com

Ranjitha.M,

computersciencedept,

IFETCOLLEGEOFENGINEERING

Villupuram,india.

Aswini.R

Assistant Professor,

IFET COLLEGE OF ENGINEERING,

Villupuram,india.

Praveen Kumar rajendran,

Cognisant Technology Solutions,

Chennai,India

Praveenkumar558@gmail.com

***Abstract*----The main concept of our survey website is to update information automatically by analyzing the twitter data.Weare using the social media as the base to survey the data and finding the trending topics which was popularly going on. The information which is widely going on in the twitter will be updated in the server as a trending topic in our surveying website. The data's in the twitter will be analyzed and by the reactions given by each and every tweet will be examined. which the sentimental analysis is used here to find the different types of reactions given in every tweet. By inspecting the reactions in the tweet, the information will be updated accordingly to it. On the authority to the number of reaction in the tweet will be arranged in the levels of trending order. The main purpose of our website is to update the information automatically without any manpower the trending topics are displayed and in the popular trending order.**

Keywords: update information automatically, analyzing twitter data, trending topics, reactions, sentimental analysis.

I. ANALYZING THE DATA

The term analysis is defined as the detailed examination of anything to understand the futures of it. Through analysis we can get clarification of an expression that has been used. Analysis can also be termed as to depict economic decisions mathematically, it can have both rational and predictable. Though analysis as a formal concept is a relatively recent development. The term analysis is the process of breaking a complex topic into smaller portion in order to get a better understanding of it. The term analysis provides resolution to any concept that can clearly understood in the regressive sense. The word analysis came from the ancient Greek word 'analysis'.'ana' means 'up','lysis' means 'loosing' therefore 'analysis' means 'loosening up'. Analysis is clearly being understood by discovering the truth of the concept.

Big data analysis-Large sets of data can be performed by collecting, organizing and analyzing to discover essential information that enclosed within the data. It is typically performed by structured, unstructured and semi-structured data. By using specialized software tools, application for predictive analysis and data mining has been performed. Initially the Big data analysis can be explained by 3V that is Volume (amount of data), Velocity (speed of data), Variety (different forms of data). Currently the speed of data is almost unimaginable. The updating of video on YouTube, sendingmail, updating tweets and other micro texts on social media usage is widely increased by enclosed large amount of data. The data can be managed, analyzed and processed via big data analysis. The amount of data will double in every year. In the past, the creation of data would have caused serious problem. Since the 3V are not sufficient to describe big data hence the additional 4V-Veracity (accuracy of data), Variability(changeability)it has the quality of being subject to variations, Visualization (mental image of data), Value (uncountable data) has been added additionally to explain the big data analysis [9]. Data mining is a well-known technology which is used in extracting and predicting hidden information from a huge database. There are many number of algorithm which has been emerged for mining patterns in which the UP-growth algorithm is one of the popular

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

algorithm in mining high utility itemset. In this algorithm, the main key constraint is the minimum node utility in producing the PHUI (Potential High Utility Itemset) in accurate. From these PHUI the high utility itemset is generated further.by this utility based mining we can reduce the complexities and has an effective and efficient on mining the data's [10].

The analysis can be applied in various fields such as Scientific analysis, Chemical analysis, Mathematical analysis, Medical analysis etc., Social network analysis is essential of studying patterns of relationship that connect social people. Thoughanalysis has many advantages it enclosed with some disadvantages like extreme time consuming, subject to increased error, difficult to automate or computerize.

II. SURVEY ON TWITTER DATA ANALYTICS

The main concept of surveying websites is to collect all trending information and publishing it to their sites by means of manpower. To overcome the situation of manpower our surveying website updates the information automatically by means of analyzing the twitter data. The process is about sentimental analysis, which it analysis each sentimental reaction on tweet and gather the reaction to update the information by means of comparing the reactions which has large counting's that reaction will be taken as the treading information in our surveying website.

III. RELATED WORKS

StephanGouws, DonaldMetzler, Congxing Cai and Eduard Hovy [1] has investigated that how the spelling, punctuation and grammar help make a user to understandable by analyzing the microblogs that inspected on twitter. The algorithm which is used in this paper is cleanser algorithm which that automatically changes or correct the words in frequent vocabulary terms but when it comes to out of vocabulary it more over fails to correct the words. The populations of different people use short forms to express their feelings which can be corrected by cleanser. The out of vocabulary terms affected by different context term in micro text.the user population can be defined based on the twitter client (web based and mobile phone based) and also by the time zone to locate the geographical location of the person to identify the context term of language that was used by that person by which the word can be determined and analyzed. We can take a set of input tokens that can be process one by one and checking out the tokens to the words in the lexicon. The context from the user bring out to different forms of expressions and different styles of lexical transformations. we can analyze the transformation types from different domains of short message that includes the extraction of letters such as walking-walkin, gone-gon, etc. and deposition of abbreviation such as doing-dng, good-gud, array-arayetc .to be the regular transformations. Thus, this paper is interested in developing population of users and context aware language processing and understanding techniques on microblogs by means of writing conventions, lexical transformation,writing conventions and noisy text cleanser

Rob Abbott, Marilyn Walker, Pranav Anand, JeanE.FoxTree, Robeson Bowmani and Joseph King [2] has proposed that all the topics which has been posted from any place and anytime and anywhere that are get analyzed. Which they analyze the utility of meta post, contextual,dependency, word based features for signaling the agreement and disagreement relation like insulting, nastiness, and emotional appeals.by this analysis they achieve the accuracies of 68%and when compared to unigram the analysis gives 5% more accuracies

In the paper "Detecting Forum Authority Claims in Online Discussions" [3] The authors are "Alex Marin, Bin Zhang, Mari Ostendorf" This paper explores the problem of detecting sentence-level forum authority claims in online discussions. Using a maximum entropy model, we explore a variety of strategies for extracting lexical features in a sparse training scenario, comparing knowledge- and data driven methods (and combinations). The augmentation of lexical features with parse context is also investigated. We find that certain markup features perform remarkably well alone, but are outperformed by data-driven selection of lexical features augmented with parse context. In multi-party discussions, language is used to establish identity, status, authority and connections with others in addition to communicating information and opinions. Automatically extracting this type of social information in language from discussions is useful for understanding group interactions and relationships.

In this paper "SENTIMENT ANALYSIS ON TWITTER DATA"[4] The authors are "Apoorv Agarwal, BoyiXie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau "They had undergone a research about analyzing the sentiment from the microblog websites. Microblog websites are nothing but sites which are used by the people to share their opinion in real time. In this research, the authors had chosen the famous microblog site "twitter". In twitter, the opinions are shared in the form of "tweets" and we classify the tweets as positive, negative, and neutral. The classification can be done by any of the three models such as feature-based model, unigram model, and tree kernel model. They have suggested that these models can also be combined to form a new model. In this paper, they had introduced new

features to classify the tweets by using hashtags and emoticons. They had introduced two new resources one is emoticon dictionary to map emoticon and acronym dictionary from the web to map acronyms. They had created an emoticon dictionary using 170 emoticons and labeled as positive, as negative and mapping acronym such as lol to laughing out loudly in acronym dictionary. The pre-processing of data has the following process replace all emoticons with corresponding sentiment polarity then replacing the words not, no, never etc as not and replacing continuity of characters into three characters for example "cooooooool" as for being taken as "cool". The data used in this experiment are 79152 tokens, 30371 stop words,23837 english words,9356 punctuation marks, 4851 capitalized words, 3371 twitter tags, 2228 exclamation marks, 942 negations, 9047 other tokens. The partial tree had been used to divide the tweets into fragments and the root node is initialized then the fragments are placed at the leaf nodes. The nodes are compared with the emoticons and acronym dictionary then the result is generated. The feature which is proposed by the Agarwal (DAL-Dictionary of Affect in Language) is used to classify the words as negative and positive. The words with less than 0.5 is considered as a negative, higher than 0.8 is considered as positive and the rest is categorized as neutral words. the combination of kernel and sentimental features model has a better performance than other combination of models. Thus, this paper concludes with there will be a future work of parsing, semantic analysis and topics modeling on twitter data.

In this paper "ANALYZING THE DYNAMIC EVALOUTION OF HASHTAGS ON TWITTER: A LANGUAGE BASED APPROACH" [5] The authors are "Evandro Cunha, Gabriel Magno, Giovanni Comarela, Virgilio Almeida, Marcos André Gonçalves, FabrícioBenevenuto" They had done a research on how hash tags are created in a twitter based on linguistics or language approach. Using the large crawled data from the twitter, authors had find that hash tags has an effective influence on the propagation of linguistics and the length of the hash tags decides the usage and understanding of the creation of successful hash tags. Hash tags help the people to create a group of people based on the certain topic and the hash tags can be created by using the (#) in front of any string. They had selected three relevant topics of that period, namely: Michael Jackson, Swine Flu, and Music Monday. Then, they built one minor short form for each one of the topics: MJ refers to Michael Jackson, SF refers Swine Flu and MM refers to Music Monday. In this paper, the authors had examined that the partition of twitter data can help us to understand about the creation of innovative hash tags based the linguistic theories. Twitter's network had been described as a directed graph. In twitter, the relations between users are not necessarily regular, this means that it is possible for someone to follow another person without being followed by him/her. It indicates that when we talk about celebrities who have millions of followers, but at the same time follow only a few users. The "rich get richer phenomenon" which is proposed by Easely and Kleinsberg which is meant by the popularity of the most common items tends to increase faster than the popularity of the less common ones. The tags which are taken for reference shows that the most popular topics are the one which is mostly used in the tweets which confirms the "rich get richer phenomenon" Thus, this paper concludes with differentiation of hash tags which spread widely from one that fails to magnetize attention. Therefore, this approach can be useful tool for the tag recommendation system for the different environments, but in future it can be used to revise the human tagging behavior.

Dong Nguyen and Carolyn P. Rose [6] has investigated the connection between active community membership and language. In an online community, discussion with participation of n-number of users traversing multiple years. The authors used text mining and machine learning methodologies to understand the connection between online communities and language use. The participant shows high emotional involvement in the discussion. The author has been applied Lave and Wenger's theory to both face-to-face and online communities. Since the authors involved one online community targeting a very specific topic failed to cover other online communities.

RamnathBalasubramanyan [7] has proposed the techniques to identify the topics within the context of the community. By using computational methods to assign sentiment polarity to learning community and blog comments that encapsulate issues tackled by blogs. The authors evaluate the use of machine learning methods to predict the emotional reactions of members to different types of news. It indicates the emotion tied to the blog and community and determine the sentiment polarity of comments in blog.

Emily M. Bender [8] has presented the AAWD corpus, and described social acts. The author explained the authority claims and positive/negative moves. An alignment move suggests the participants agreeing and disagreeing with other participants regarding a particular topic. By comparing the widespread of authority claims and includes discussions annotated for alignment moves of social acts. The author focused on his future work to explore the range of variation among the linguistic and alignment categories across communication

media as well as the possible roles of additional categories in social acts and beyond.

IV. OUR PROPOSAL

With the reference to the paper survived, we are proposing a model to predict the user emotion using the twitter data. The proposed model of survey on twitter data analysis will be implemented using HTML, CSS as front end and python framework as back end to analyze the data. The tweets can be analyzed and characterized based on the emotions used by the social users. The twitter data analysis updates the tweets information automatically by means of analyzing the twitter data. It analysis each sentimental analysis which enclosed on twitter data. Our model is going to work based on the twitter media by analyzing the data's and interpret the sentimental emotions to update the information on our website. which this analysis on twitter data will be done automatically by means of inspecting the emotions on each tweet and updates the information of trending data in a trending order automatically in our survey website. This above process is achieved with help of the tools which the main back end process we are going to use in the process is python framework to update the trending topics and information automatically in our website.

V. FUTURE SCOPE

The above proposed model of Survey on twitter data analysis is for social users, where the number of users posted their tweets on twitter, it can be utilized by surveying the tweets based on the emotions. It automatically updated on websites to know what's the recent familiar discussion enclosed by the social users. The above proposed works monitors all the tweets on the twitter and updated it on the sites automatically, it can be implemented in very low cost and within a short span of time. The future scope of the paper wouldbe implementing the Twitter data analysis on twitter and updated it on websites automatically. The implementation of our website is done by some of the tools, the website designing will be done with HTML, CSS and some of the web development tools for interaction between the twitter and our survey website. For analyzing the data in the twitter and for sentimental analysis for each and every tweet to inspect the emotions given by the twitter user will be processed as back end that we are goinng to use the python framework to analyze the data and updating the information automatically in our web site. The main advantage of our model which is better than other model is to updating the trending topics and information automatically in our website without any manpower.

References

[1].Stephan Gouws, Donald Metzler, Congxing Cai and Eduard Hovy "Contextual Bearing on Linguistic Variation in Social Media" workshop on language in social media (LSM 2011), page 20-29(23 june2011)

[2].Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani and Joseph King "How can you say such things?!?: Recognizing Disagreement in Informal Political Argument" workshop on language in social media (LSM 2011), page 2-11(23 june2011).

[3].Alex Marin, Bin Zhang and Mari Ostendorf "Detecting Forum Authority Claims in Online Discussions"workshop on language in social media (LSM 2011), page 39-47(23 june2011).

[4].Apoorv Agarwal, BoyiXie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau "Sentiment Analysis of Twitter Data"workshop on language in social media (LSM 2011), page 30-38(23 june2011).

[5].EvandroCunha, GabrielMagno, GiovanniComarela, VirgilioAlmeida, MarcosAndr´eGonc¸alves and FabricioBenevenut "Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-Based Approach" workshop on language in social media (LSM 2011), page 58-65(23 june2011).

[6].Dong Nguyen and Carolyn P. Rose "Language use as a reflection of socialization in online communities" workshop on language in social media (LSM 2011), page 76-85(23 june2011).

[7].RamnathBalasubramanyan, William W. Cohen, Doug Pierce and David P. Redlawsk "What pushes their buttons? Predicting comment polarity from the content of political blog posts" workshop on language in social media (LSM 2011), page 12-19(23 june2011).

[8].Emily M. Bender, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson, Alex Marin, Bin Zhang and Mari Ostendor "Annotating Social Acts: Authority Claims and Alignment Moves in Wikipedia Talk Pages" workshop on language in social media (LSM 2011), page 48-57(23 june2011).

[9].PraveenKumarRajendran,A.Asbern,K.ManojKumar,M.Rajesh, R.Abhilash."ImplementationandanalysisofMapReduceonbiomedicalbigdata."IndianJournalofScienceandTechnology9.31(2016).

[10].A.PioSajin,R.Ashwini,B.BaronSam."DistributedandImprovisedUp-GrowthApproachforUtilityBasedMining"ResearchIndiaPublications(2015)