User Classification in Twitter Based on Sentiment Analysis on user Tweets

Aarabhi Babu[1] and Vince Paul[2]

[1]UG Scholar, [2] Professor,
[1,2]Department of Computer Science and Engineering,
[1,2]Sahrdaya College of Engineering & Technology, Kodakara,Kerala, India
aarabhi.babu@gmail.com[1],vinceakkara@gmail.com[2]

Abstract : This work user classification in twitter deals with classification of the twitter users based on their tweets till dated. Each tweets and re tweets of each users are taken and sentiment analysis is performed .By aggregating the sentiment score of whole tweets of the user his total user score is calculated based on the user score users are classified into different groups like highly positive neutral positive highly negative neutral negative .users who came under negative side will be warned first about their negative behaviour and if they continue they will be blocked.

Keywords: Tweets, Sentiment Analysis, user classification, text emotions, opinion mining, user score

I. Introduction

Sentiment analysis sometimes known as opinion mining or emotion AI refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation appraisal theory, affective state that is to say, the emotional state of the author or speaker, or the intended emotional communication Social networking sites are of the greatest influential medium among public beyond age limit. People share their knowledge, feelings, opinions views etc freely through these sites. social networking site like twitter is a great platform where the contemporary issues are discussed within hours and these create a large impact on public on making their views therefore there is a great necessity that these sites should be continuously monitored and remove malicious and nuisance tweets. It is been widely reported that people are misusing these sites as a platform to speak out whatever they want and to spread anti social thoughts among public .Thus a monitoring system over these sites are very essential, This work comes out with such an idea that a monitor system over twitter and keeping a track over the tweets and the users, calculate sentiment analysis score of each tweet and public post till dated and then aggregating the whole score of each tweet then user score is calculated . Based on the user score they are classified into different groups. Score for each post is recalculated based on the sentiment analysis performed on its comments.

International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.5 (2017): Page.1735-1740
www.ijcsec.com

II.   Related Work

There are basically two approaches for performing Sentiment analysis. One is using the word lexicons like SentiWordNet2, OpinionFinder3, and the other uses machine learning techniques. Andrea Esuli describes SentiWord- Net in[3] as a publicly available lexical resource for opinion mining. The resource has sentiment scores to various words that roots from WordNet4. Alaa Hamouda and Mohamed Rohaim[4] proposed and implemented a technique for product review classifcation which uses SentiWordNet to classify reviews into different sentiment categories. Similar classification of user reviews was done by L Banic and A Mihanovic [3] on reviews collected from the web about different hotels. Barbosa and Feng classifed tweets into positive and negative sentiment entities, by employing the classification techniques of the Naive Bayes(NB), Maximum Entropy(MaxEnt), and the Support Vector Machine(SVM) [5]. These are the major machine learning classifiers, and their results show that maximum accuracy is provided by NB and MaxEnt. Brendon O Connor et al [6] follows sentiment analysis of political opinions, mined from Twitter posts, to automatically summarize the results of opinion polls, prior to a chosen political election event. While Jakub Pikorski[7] exploited the micro blogging site in order to predict border-security in European countries, Antonie Boutet et al.[8] used the followers to following ratio on twitter to predict the UK General Elections 2010. Most feature oriented sentiment analysis techniques consider adjectives, adverbs and nouns as features while Mostafa and Ali (2012)[9] used an approach that takes up verbs as the major feature for sentiment extraction. Most of the existing sentiment classification models work very well for structured, long documents or paragraphs. Models intended for social networking sites need more optimization since data from such sources are often shorter, and not in proper language in terms of grammar and spellings. This is a growing area of research. [5] [9] Most of the published literature are on data mining that classifies user reviews about products or movies and uses either of the two approaches of sentiment analysis. There is a need to develop a model that extends sentiment analysis to track the origin and measure the influence of malicious posts on websites like Twitter. This may require additional parameters like, a measure of influence of the posts, detailed examination of the originating user accounts and a combination of sentiment analysis approaches.

III.   Proposed Work

According to the algorithm each word from the given tweet is extracted and will check with the training data set if the extracted word is present in the data set then its corresponding sentiment score is extracted and stored and when it reaches the at the end of line all the sentiment score will be added and along with the count of each words in the sentence will be taken that is , total number of words in the sentence once the total count and total sentiment score of the given tweet is calculated as,

TOTAL SENTIMENT SCORE AND TOTAL COUNT

If the given sentence contain any words such as 'not','very' it is treated as special cases. During analysis each tweet and word will be extracted and checked to not , if match occurs ,we will check for the next word and if the next word is positive word or a negative word the sentiment score of next word will be multiplied with negative one and will be added to the sentiment score. If the match occurs for the word very sentiment score of very next word is extracted and check for positive or negative . if the very next word is positive then a score of+2 will be added instead if the next word is negative a score of-2 is added. For the calculation of user score total score of each user has to be calculated, for that first while calculating the sentiment score of his post after calculating the sentiment score value of his post in order to determine its positive post or negative post we will check the sentiment of its comments too.

International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.5 (2017): Page.1735-1740
www.ijcsec.com

If the comments of the tweets are more than threshold value say 10 then each tweet is evaluated sentimentally the sentiment score and count of re tweets are then considered for the determination of value of main post. positive score and total positive count of comments and negative score and total negative comments are counted. then positive count and positive score is multiplied and same in case of negative. Then the net result is added to the sentiment score of main tweet.

Once user score is calculated users are classified based on the score of total sentiment score by total number of tweets the tweeted

$$\frac{\text{Sentiment score}}{\text{Total number of tweets}}$$

ALGORITHM

Input:set of tweets T,database D

score =0
For each token t in T
If(t='very')
Fetch next word
S<- sentiment score of next word
If(s>0)
S=S+2
Else if (S<0)
S=S-2
End if)
If(t != 'not'
S<-Sentiment score of each token
 score =score+s

End if
    If token='not'
S<- sentiment score of not
Score=score+s
Fetch next word
S<- sentiment score of next word
End if
S=s*-1
Score= score+s

Calculating sentiment for main posts

Calculate SA for tweets
sentitweets()
p.score=0
n.score=0
p.count=0
n.count=0
s.score=0

International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.5 (2017): Page.1735-1740
www.ijcsec.com

If
 c.tweets>10
  s<- sentiment of tweet
S>0
p.score=p.score+s
p.count=p.count+1
S<0
n.score=n.score+s
n.count=n.count+1
A=p.score*p.count
B=n.score*n.count
C=A-B
X=2(priority factor)
Postsenti=(x*sentiscore)+C+L
Userscore = $\sum$Postsenti

IV.   Experiment Results

The above figures shows different windows of the work figure a denote login page where the user can register and login figure b shows the home page of user where the user tweet and watch other posts. figure c is the window of a blocked user. Analysis of the work include 20 different sentences were given to 10 different people and ask them to calculate its sentiment score and the same sentence where given in the work and calculate its score and match between these two were calculated and a graph is plotted based on the match
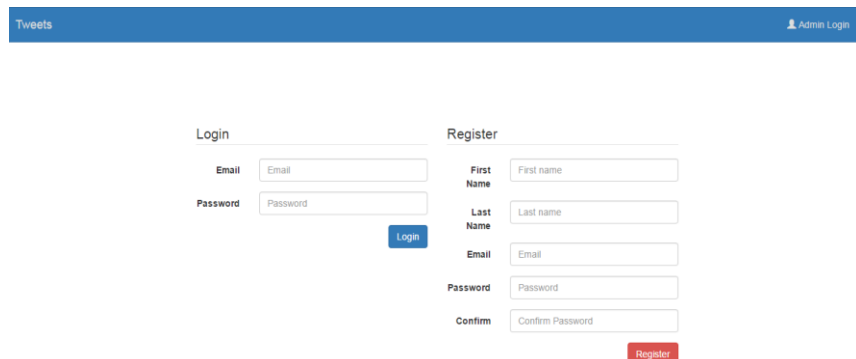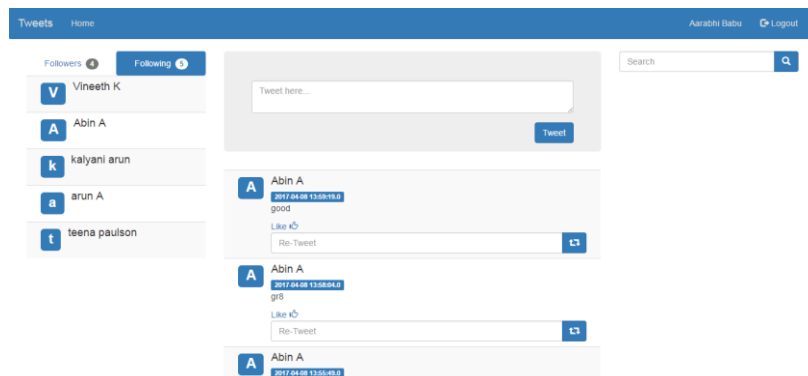

Figure. (a)

Figure.(b)

International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.5 (2017): Page.1735-1740
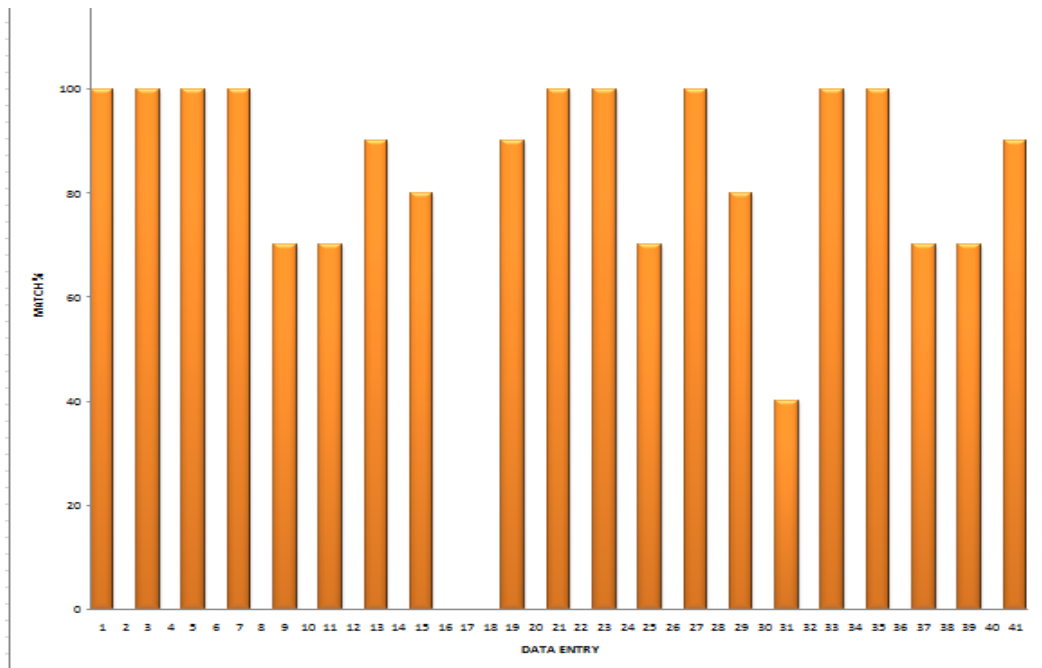www.ijcsec.com

Figure.(c)

Figure.(d)

V.   Conclusion

We have Created the twitter based user interface where the user can tweet, re tweet and follow other users. Algorithm for the sentimental analysis for the tweets is implemented Special cases like not, very and short words are implemented. Calculated User score based on their tweets till dated Classified the users based on their score and applied warning, blocking functionality.

REFERENCES

[1] Shilpa P, Madhu Kumar S D. Feature Oriented Sentiment Analysis in Social Networking Sites to Track Malicious Campaigners", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) 10-12 December, 2015

[2] Rui Xia,FengXu, Zong, and Qianmu Li. Dual Sentiment Analysis: Considering   Two Sidesof One Review . IEEE TRANSACTIONS 2015.

[3] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of 5th Conference onLanguage Resources and Evaluation, 78, pages 417-422. AIAA, 2006.

[4] Alaa Hamouda and Mohamed Rohaim. Reviews Classification Using Sentiwordnet Lexicon. The Online Journal on Computer Science and Information, 2:492 -524, 1997.

[5] Luciano Barbosa and Junlan Feng. Robust Sentiment Detection on Twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics,pages 36{44, 2010.

[6] Brendon O'Connor and Balasubramanyan From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In InternationalConference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Feb 2014.

[7] J Piskorski and Frontex. Exploiting Twitter for Border Security-Related Intelligence Gathering. In IEEE European Intelligence and Security Informatics Conference (EISIC), pages 239 -246 Aug 2013.

[8] Antounie Boutet et al. What's in your Tweet?:I know who you supported in uk 2010 General Elections. In IEEE European Intelligence and Security Informatics Conference (EISIC), 2012.

[9] M Karamibekr and Ghorbani A.Ali. Verb Oriented Sentiment Classification. In
IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), pages 327 - 331, May 2012.

[10] Efthymios Kouloumpis,Theresa Wilson,Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG,Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011

[11] Ashish Shukla, Rahul Mishra. Sentiment Classification and Analysis Using Modifed K-Means and Nave Bayes Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering , 2015.

[12]. Alessia DAndrea,Fernando Ferri,Patrizia Grifoni ,Tiziana Guzzo Approaches, Tools andApplications for Sentiment Analysis Implementation ,International Journal of ComputerApplications (0975 8887) Volume 125 No.3, September , 2015.

[13] Bo Pang and Lillian Lee .Opinion mining and sentiment analysis , Foundations and Trends in Information Retrieval , 2008.

[14] M. Gjoka, M. Sirivianos, A. Markopoulou, and X. Yang.Poking facebook: characterization of osn applications." In Proceedings of the first workshop on Online social networks, WOSN,2008.

[15] Zohreh Madhoushi,Abdul Razak Hamdan,Suhaila Zainudin. Sentiment Analysis Techniques in Recent Works,In Proceedings of the first workshop on Online social networks, 2015.

[16] Amna Asmi,Tanko Ishaya. Negation Identifcation and Calculation in Sentiment Analysis, The Second International Conference on Advances in Information Mining and Management, 2012.

[17] Yahya Eru Cakra, Bayu Distiawan Trisedya. Stock Price Prediction using Linear Regression based on Sentiment Analysis , ICACSIS , 2015