Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4

2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 978-1-6654-1960-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICV50876.2021.9388596

People's Behaviour Analysis in Chat Message using Natural Language Processing

[1]V.Selina Annie Retna, [2] Prof. P. Brundha [3]Dr. RajKumar G

[1] PG Student, Francis Xavier Engineering College, Department of CSE
[2]Head of the Department, Francis Xavier Engineering College, Department of CSE
[3]Professor, Francis Xavier Engineering College, Department of ECE
Tirunelveli, India.
[1]selinaannieretna@gmail.com [2] brundha@francisxavier.ac.in [3]gmanly12@gmail.com

Abstract

Nowadays, the mode of communication is mainly through messages. A lot of information has been conveyed through WhatsApp. WhatsApp is the most popular chat application with active users of more than 650 million. It has been widely used by all, especially among the business people and youngsters. Using several analyzing tools, users can analyse the WhatsApp group chat or personal chat. Authentically users wish to analyse their chat for several purposes. This research work is intended to perform a flirt analysis and time analysis. This project has many use cases like the parent, who wants to analyze their child chat; the police, who want to get valuable information from culprit chat; the business people, who wants to know the status of the business in the group chat. Using the Deep Learning model (NLP), sentimental analysis has been performed for each text. This helps to find the state of mind of the chatters. Further, this research work calculates the number of positive and negative statements that are used by each person in the text by using the text mining concept. As now due to this pandemic situation, every conversation and also the important discussion has been done through the WhatsApp and it was highly needed for the person who wants to check their child's conversation and also for the higher authority for enquiry and for the business chair person who are needed to analyse their business well being group can also be used for their personal usage of analyse using the algorithm in this method.

Keywords: Sentimental Analysis, Data mining, Emotional, Natural Language Processing

INTRODUCTION

Nowadays, the usage of social media networks now became a common mode of information sharing. A large set of users were now adapting to this model of the technological era. The usage of social media now became common for sharing the messages the video and also the pictures and not only for the personal purpose it was also used for the professional purpose as sharing or advertising the business-related details also nowadays became the new normal.

For the privacy and safety of the users, as of now mostly our chat conversation is through WhatsApp and by the sentimental analyze model we can analyze and we can conclude whether the chat we are going safe or going correctly and also the parents can track their child move and also the business personality can also check how they got their review in the group works.

As now all were going on through the data billions and trillions of data are passing every day and also for the analysis and to enhance the performance and detect the performance of the sentimental analyzing methodology, which was introduced for achieving the better result regarding the business and also regarding the privacy it will be very much useful for the detection.

The text mining concept is also used for mining the text data, which was used in the WhatsApp conversation.

LITERATURE SURVEY

As it is already known that, the sentiment part was playing a major role in one aspect of life. It also has a major impact on one's life. Sentimental analysis part was done in the Twitter as now most of them share their views on the Twitter and also searching their procedures of public opinion over there and also by introducing the sentimental analysis part of the geo tweets, it was made to place whether the process makes a positive impact or not. With the spatial and the temporal methods, the dataset was

analyzed and regarding that the sentimental analysis will be calculated [8].

Most commonly for calculating the data which was focused on the particular topics of the social media, [9, 13] but in the recent days of research and the most validation was conducted to conclude the sentimental analysis.

For every process of calculation, the data will be collected and then for the pre-processing technique, the data collected was given to certain classifying techniques for performing the sentimental analysis [13].

This method analyzes the tweets, which were given by various persons by collecting the data and thus made them to the processing stage to obtain the results of the analysis [2].

The sentimental analysis was the main factor in the social media marketing in order to which they can be used for their growth in their respective business by calculating their comments based on the specific product or the course, by which they come to the specific conclusion whether they are in the positive or the negative approach or with the neutral set of reviews and they made their method of approach.

The sentimental analysis in Twitter using the ordinal regression for analyzing and revising for the respective data which was made with the execution procedures. The proposing of the human-centered approach of maintaining the details with that of equally maintaining with the required standard of obtaining the emotional information from human and technical sensors to get those values. The tweets were most probably obtained from the dataset which we were taken along with the tweets and which were posted down by the local residential users [14, 1].

Some sentimental analysis regarding that of the process of giving the feedback for the Facebook game review system was introduced and based on that of the sentimental analysis results we can come to know about the analysis report whether the views were positive and loved by everyone or else negative or they are neutral from which they will conclude whether it was liked by everyone or not [7, 18].

Clustering of the data values will result in the efficient analyzing content with that of required challenges and also relatively with the further challenges and hopefully, the clustering algorithm was mainly used in order which that was used for the grouping of the unstructured and the structured data with the relevant data procedure and they will group them relatively with the efficient continuous procedural structures which will be used for the analyzing of the sentimental process.

Density-based clustering techniques were used in the analysis for the sentimental analysis in the required dataset which was searched and was taken from that of the Twitter data values. In which the error-based zone was done with that of the relevant data which was most commonly represented with the word-based analyzing techniques. Segmentation of this process will be done by that of the stickiness score which was provided to them by the global and the local users [24, 26].

The individual emotions were taken for the analyzing purpose and then it was made for the test and by which they get the information across the Twitter and the required response was noted.

Social media was the main mode that relies on their preferences and their prediction methodologies. They mostly rely on the preferences which were made by them in which they can be collected and were analyzed and was certainly included with the related method conclusion structures. The preferences for anything related to education, business, Movie, work, and any other things will be subjected and was analyzed, and based on their preference prediction it will be classified and was sorted out for the required structure [4, 23].

The framework of the reduction was mainly based on the regression for the classification and it can be available to perform the classifier method of regression structure thereby maintaining three types of the steps followed for organizing learning and classifying the rules which were implemented in it [14].

Identification of the sentiments now a day was the most needed issue and a problem and this was used for automatically finding their opinions and based on them. It was analyzed and classified by combined with supervised learning.

Proposed System

The system focuses on the extraction of the WhatsApp chat and was considered for the sentimental analysis using the deep learning model i.e. the natural language processing by following some steps.

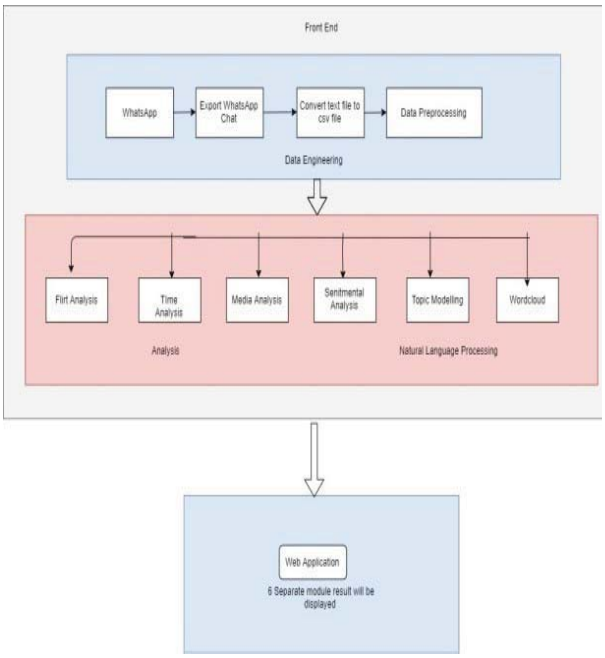

Fig: 1 Block diagram

Methodology

Step1: Data collection

Data collection is the step in which we are going to collect the data from various users having the chat conversation with their friends and known people. WhatsApp has an option to export our chats as a text file. An exported text file will be converted into a CSV file for easy analysis.

By this we can collect the data and can store the details for the sentimental analysis.

Step2: Data Pre-Processing

Collected data will be made pre-processed which means encoding the categorical information in the data, dropping unwanted parameters, scaling the parameter values to achieve normal distribution (Zero mean and Standard Deviation as one), handling missing values, and so on.

Pre processing is the method in which the data will be undergone into certain steps which can be made for the future analysis procedures which will be efficient for the calculation procedures

The preprocessing steps which were taken place in the process were:

- Step1: Classification of the date, chat, time, text, name.
- Step 2: Processing the emoji free chat was made.
- Step 3: Making the emoji deemojifying in the preprocessing steps.
- Step 4: The excessive gaps in the documents were analyzed and removed.

Step3: Flirt Analysis

Flirt Analysis is the step in which the major part of the analysis was done. The main aim of the module is to detect the flirt which was done between the peoples via chat by comparing the chat with the list.

In the flirt analysis, the text message will be compared with the flirt list. The total number of flirt words used by a person will be divided by the total number of flirt words. The flirt percentage will be calculated for each person in the chat.

Flirt Analysis Algorithm

Step 1: In a flirt analysis, first we have to collect the flirt words.

Step 2: Segregate the message respective to the chatter's name.

Step 3: Calculate how much unique word used by each chatter (U)

Step 4: Calculate total words used in each row (T_w)

Step 5: Calculate unique frequency for each word (U_w)

Step 6: Formula: Unique Frequency for a Word (U_w)=(U/sum of T_w)* 100

Step 7: Calculate the total number of flirt list (F_l)

Step 8: Check each word with a collected flirt list

Step 9: Store the used flirt word by each person as a separate list(FW)

Step 10: Calculate the total number of flirt words used by each person. (L_f)

Step 11: Calculate the Unique Flirt frequency of each word for each person (F_f)

Step 12: Formula for Unique Flirt frequency (F_f) = (U/F_l)*100

Step 13: Calculate Total Flirt Percentage=sum of F_F/ total number of F_w

Step4: Time Analysis

In time Analysis, the date and time of each text will be separated, can able to predict the most active date, the most active hours, the active day, and average message per day. This kind of parameter will be helpful for those who want to peak the date and time of the chat.

Max (Repeated date, Date, Time)

Step5: Media and Call Analysis

WhatsApp has the option to send photos, voice notes, videos, and audio to anyone. This type of option has both benefits and disadvantages. So here we can calculate how many Media are shared by both the person. In addition to this WhatsApp has an option for deleting the message, voice, and Video call, anyone can easily connect with people without more struggle. We can able to find how many calls (Missed Voice and Video call) are tried by both the person and how many messages are deleted

Media shared count [Query (Select the media shared)for each person in the chat]

Step 6: Topic Modeling:

In the topic modeling, the natural language processing will analyze the entire chat document and thus comes up with a suitable topic for the entire document according to the respective state of chat we produce.

Step 7: Word Cloud:

Word Cloud is a visual treat of text words that are the most frequent word used by the chatters in the order of increasing font size too small font. This helps us to see the most frequent word.

Step 8: Django framework

All the mentioned modules are displayed are processed and displayed using Django (python) as Web Application

Performance Analysis:

1. The project was made to undergone the sentimental flirt analysis in the private setting of a group in the WhatsApp chat and it was implemented and was given out the exact result for the analysis
2. The working process was undergone with first starting with the pre processing steps and after that, it basically checks the input with the flirt list and thus processes and then the result was to be displayed. Exactly what was the expectation from this flirt analysis methodology was obtained in this analyzing process
3. The procedures were created so that they can logically perform the sub-functions for the flirt analysis technology
4. All the codes which were given for obtaining the result were readable and it can be easily taken for the understanding and the structure of the progress all was implemented according to the procedures

Conclusion and Future work

In this paper the sentimental flirt analysis methods was implemented in the all the entire chats and was given out the respective result can be obtained

The entire methodology can be implemented but for the storage the cost was little high for using the cloud storage and can be reduced in the future works

The respective chat messages can be taken out from various users with the approval of the respective users it was having some delay in execution can be used effectively in future works.

REFERENCES

[1]"Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers",Gonzalo A.Ruz[ab]Pablo A.Henríquez[a]AldoMascareño-8 January 2020

[2] Mohammad A.Hassonah, RizikAl-Sayyed, AliRodan, Ala' M.Al-Zoubi[c]IbrahimAljarah[a]HossamFaris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter"- 11 December 2019

[3]Shikha Tiwari; Anshika Verma; Peeyush Garg; Deepika Bansal,"Social Media Sentiment Analysis On Twitter Datasets"

23 April 2020 -**ISSN:** 2575-7288"Sentiment analysis using deep learning architectures: a review" Ashima Yadav & Dinesh Kumar Vishwakarma- 2 dec- 2020

[4] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in mining text data: Springer, 2012, pp. 415-463.

[5] Y. Wang, "Sensing Human Sentiment via Social Media Images: Methodologies and Applications," Arizona State University, 2018.

[6]S. D. Pressman, M. W. Gallagher, and S. J. Lopez, "Is the emotion-health connection a "first-world problem"?," Psychological science, vol. 24, no. 4, pp. 544-549, 2013

[7]P. Garg, H. Garg, and V. Ranga, "Sentiment analysis of the Uri terror attack using Twitter," in 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017: IEEE, pp. 17-20.

[8] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series.," Icwsm, vol. 11, no. 122–129, pp. 1–2, 2010.

[9] M. A. Cabanlit and K. J. Espinosa, "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons," in Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on, 2014, pp. 94–97.

[10] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 1367.

[11] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 625–631.

[12] H. Saif, M. Fernandez, Y. He, and H. Alani, "Evaluation datasets for Twitter sentiment analysis," Emot. Sentiment. Soc. Expressive Media, p. 9, 2013.

[13] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," in 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 628–632.

[14] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," Processing, vol. 150, no. 12, pp. 1–6, 2009.

[15] V. Singh and S. K. Dubey, "Opinion mining and analysis: A literature review," in 2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 232–239.

[16] W. Chu and S. S. Keerthi, "Support vector ordinal regression," Neural Comput., vol. 19, no. 3, pp. 792–815, 2007.

[17] S. Liu, F. Li, F. Li, X. Cheng, and H. Shen, "Adaptive co-training SVM for sentiment classification on tweets," in Proceedings of the 22nd ACM international conference on Information & Knowledge Management, 2013, pp. 2079–2088

[18] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data: Springer, 2012, pp. 415-463.

[19] Y. Wang, "Sensing Human Sentiment via Social Media Images: Methodologies and Applications," Arizona State University, 2018.

[20]S. D. Pressman, M. W. Gallagher, and S. J. Lopez, "Is the emotion-health connection a "first-world problem"?," Psychological science, vol. 24, no. 4, pp. 544-549, 2013.

[21] M. E. Geisser, R. S. Roth, M. E. Theisen, M. E. Robinson, and J. L. Riley III, "Negative affect, self-report of depressive symptoms, and clinical depression: relation to the experience of chronic pain," The Clinical journal of pain, vol. 16, no. 2, pp. 110-120, 2000.

[22] S. D. Pressman and S. Cohen, "Does positive affect influence health?," Psychological Bulletin, vol. 131, no. 6, p. 925, 2005.

[23] S. Gohil, S. Vuik, and A. Darzi, "Sentiment analysis of health care tweets: a review of the methods used," JMIR public health and surveillance, vol. 4, no. 2, p. e43, 2018.

[24] L. Anselin, "Local indicators of spatial association—LISA," Geographical Analysis, vol. 27, no. 2, pp. 93-115, 1995.

[25] R. M. Assuncao and E. A. Reis, "A new proposal to adjust Moran's I for population density," Statistics in medicine, vol. 18, no. 16, pp. 2147-2162, 1999.

[26] Y. Fan, X. Zhu, B. She, W. Guo, and T. Guo, "Network-constrained Spatio-temporal clustering analysis of traffic collisions in Jianghan District of Wuhan, China," PLoS One, vol. 13, no. 4, p. e0195093, 2018.

[27] M. j. Fortin, M. R. Dale, and J. M. Ver Hoef, "Spatial analysis in ecology," Encyclopedia of environmetrics, vol. 5, 2006.

[28] J. D. Morenoff, R. J. Sampson, and S. W. Raudenbush, "Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence," Criminology, vol. 39, no. 3, pp. 517-558, 2001.

[29] L. A. Waller and C. A. Gotway, Applied spatial statistics for public health data. John Wiley & Sons, 2004.

[30] S. J. Rey and B. D. Montouri, "US regional income convergence: a spatial econometric perspective," Regional studies, vol. 33, no. 2, pp. 143-156, 1999.

[31] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora," in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, 2009: Association for Computational Linguistics, pp. 248-256.

[32]B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series.," Icwsm, vol. 11, no. 122–129, pp. 1–2, 2010.

[33] M. A. Cabanlit and K. J. Espinosa, "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons," in Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on, 2014, pp. 94–97.

[34] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 1367.

[35] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 625–631.

[36] H. Saif, M. Fernandez, Y. He, and H. Alani, "Evaluation datasets for twitter sentiment analysis," Emot. Sentiment. Soc. Expressive Media, p. 9, 2013.

[37] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," in 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 628–632.

[38] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," Processing, vol. 150, no. 12, pp. 1–6, 2009.

[39]N. R. Kasture and P. B. Bhilare, "An Approach for Sentiment analysis on social networking sites," in 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 390–395.

[40] V. Singh and S. K. Dubey, "Opinion mining and analysis: A literature review," in 2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 232–239.