Analysing and Predicting the Emotion of WhatsApp Chats Using Sentiment Analysis

5 authors, including:

Bharath Sai Reddy Chinthapanti

VIT University

4 PUBLICATIONS   **3** CITATIONS

SEE PROFILE

Kowshik Sola

VIT University

3 PUBLICATIONS   **3** CITATIONS

SEE PROFILE

Nikhil Kumar Reddy Obili

VIT University

2 PUBLICATIONS   **1** CITATION

SEE PROFILE

Gopichand Gi

VIT University

9 PUBLICATIONS   **20** CITATIONS

SEE PROFILE

Analysing and Predicting the Emotion of WhatsApp Chats Using Sentiment Analysis

Chinthapanti Bharath Sai Reddy 1, Kowshik S 2, Rakesh Kumar M V 3, O Nikhil Kumar Reddy 4, Gopichand G 5*

1,2,3,4,5School of Computer Science and Engineering,Vellore Institute of Technology, Vellore, India.

Email: gopichand.g@vit.ac.in

Abstract

What does other person feel about us is the most frequently self-asked question, Everyone has the curiosity of what other person thinks about the other while having a conversation ,judging the other person can't be done perfectly ,So this paper is providing a way using sentiment analysis between conversation .While chatting with other person we always have a question about our image on the other persons mind. This process deals with pre-processing the data obtained from the WhatsApp chat which is exported to a server and then sentiment analysis is applied for each message and all of the messages' sentiment is normalized from a proposed method and overall sentiment is found out.

Keywords:WhatsApp,Classification algorithms, Chats, LSTM, Bag of words,Machine learning, Sentiment analysis

Introduction:

Curiosity is one of the major thing which pushed many people to do many astonishing things also same curiosity kills person internally, Everyone have the curiosity about their image in the other person's mind based on the conversation that took place between them, also the person mood and emotion will be based only on the past conversations as well.

So the solution this paper presents is applying sentiment analysis to the messages that are exported from the chat.

All the messages are exported from WhatsApp export feature available in the app itself. This exported chat has to be pre-processed for better computation and all the messages are separated and sentiment analysis is applied on each of them.

Sentiment analysis is a method or a sub process of natural language processing and data mining that is basically used for the identification of users opinion or emotion. Also this concept works on the core principles of machine learning where a data set(corpus in terms of NLP) is used to train and process that data and a model

is generated and used for evaluation of emotion of test chat. This process has evolved into a very powerful field in text analysis.[10]

There are two methods in getting the result

Using statistics :

In this method all the algorithms treat messages as bag of words in which order of the words. The main or the old text is processed to a bag of words where the emotion is predicted. Here all the stop words are removed and just sentiment is applied on the remaining words which are assumed to have the sentiment.

Using Statistics and linguistics

This is a method where both natural language processing and statistics is used in a hybrid manner in order to predict the emotion.

Group analysis has been made by [1] and the results are dismal and most important conclusions cannot be drawn out of it. This paper has explored the concepts of Sentiment Analysis and Data Analytics. Also having a huge insight in the knowledge sources in the recent development of the sentiment analysis still lexicons are most used in almost all of the well-known algorithms. The ideas can be additionally investigated to know more issues and difficulties in the region of Sentiment Analysis. The examination can be additionally reached out to Machine Learning ways to deal with characterize the content and make expectations through Supervised or Unsupervised Learning.[2]In this paper sentiments in visits had been ordered into 10 distinct classifications to examine the direction of the gathering. The content was separated utilizing devices and afterward parsed to yield most fitting words. Classification of

content was finished by contrasting them and lexicons in installed libraries.For analysing sarcastic statements they have to be manually removed from the unrelated context and processed separately for every attribute.

Classification of sarcastic sentences is a complex task as they have to be processed physically for every attribute. Resultportrays the weightage of sarcasm identification whatever may be the level of sarcasm and also what ever the context was meant. From the results, we can see the estimation of an announcement containing mockery sentiment. This element can be utilized to identify mockery sentiment from the WhatsApp bunch with Indonesian language content.[3]

Finally, this paper results show that users consider the emotional expression of all kind of sentiment as valid on social media. Positive and negative emotions doesn't affect our overall predication where we are giving opposite sign weights for them which we will be discussing in the proposed work section. The comparatively higherscaling of positive emotion expression appropriateness points toward a stronger relevance of positivity norms, which corresponds with positive authenticity expectations on social media . But at any cost we should consider these results as legitimate and one of a kind to online features. For the most part, the way that individuals adhere to rules of cooperation comes from the natural need to dodge the danger of social authorizations and dismissal.In [5], results showedtrendminer a method for real time analysing understanding ofWhatsApp texts which is an open source tool to operate on text analysis tasks on WhatsApp data.

The study in this paper analyses the messages and shows that these messages are the main source of contents which shows the way to cyberpsycho which is an effect discussed in this paper.The examination right now the instant messages and shows that these messages are the potential substance which prompts cyberpsycho effect. The work right now the strategy to analysis the volume of sentiments in the substance sent or spread utilizing well known web based life basically WhatsApp to break down cyberpsycho assaults; feelings and sentiments analysis, thoroughly. The results displaysalmost many of the text messages contain emotions so this will pivot in the direction to cyberpsychois mostly used when a context is well know in the fields os sports, politics ,social games and economical situation . [6]

Problem Description:

In this tech filled world most of the time people spend talking to others in text format and specifically in WhatsApp ,since it is the number one messaging app in the world with almost 1 Billion active users and also with a strong encryption which works on the model of end to end encryption. How other person feels towards us while having a conversation is one of the biggest questions even most trained professionals cannot answer it. This paper deals with the creating an automation process for identifying overall sentiment of the individual person. Also the frequency of the messages by each person also affects the person emotion. There is no third part application that does this works and also there lies more benefits from this process for both individual and also corporate offices for analysing the employee performance and their activeness in their project groups.

Proposed Work:

Dataset is a simple text file that has been extracted from any of the WhatsApp groups or one to one individual chat. More the number of text messages ,the more the accuracy will be in identifying the sentiment. Chat from the WhatsApp can be extracted using a feature called export chat and this will mail the compressed zip file which has a text file of the chat from the beginning and all the undeleted chat will be included in this text file. A lot of pre-processing needs to be done. In this paper we used a algorithm named LSTM(long short term memory).

To explain what LSTM is ,here is a general way to understand it. We homo sapiens don'tintiateour thinking from tyro level every second. As we start skimming through any essay, This method learns each word based on the past sentences that have been present in the corpus. They don't simply leaveand start from scratch as normal CNN's do.

Old neural networks such as CNNs cant do the memory storage of previous context. For any scenario in real life past experience and context is must to participate in the present scenario.

RNN has the solution for this. They are work with networks having loops in them, by providing a solution for long lasting of data.

LSTM Networks

Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber. Their performance in the huge and real world datasets is astonishing and used everywhere in large scale projects.

LSTMs are mainly used and are made to explicitly stop the long-term dependency problem. Storing the memory for a long time is their inherit property which makes them stand out from all other neural networks. Use of traditional neural networks for this kind results in the statics classification of each message which doesn't provide more information about the conversation in the chat that we are inserting for the processing and prediction.

We are using a pretrained LSTM model which was trained on large datasets of social media and also online social news platforms of different sources withholding different corpora.

Text Pre-processing needs to be done as most of the chat will have emojis and media. Media can be eliminated while exporting the chat, WhatsApp alerts with a pop up displaying whether to export with media or without media, processing with media is a whole new level of complex analysis that's for another day. Here the text messages have metadata attached to it.The time of the message that was send and also the sender name followed to it .Also we have tested with various categories and the ones that are encoded with the LSTM model that was used are

- Angry
- Sad
- Fear
- Happy
- Surprised
- Frustrated
- Excited

These are the basic categories which the message will fall into and each time different priority is given to each emotion and then overall emotion is tested.

Architecture :
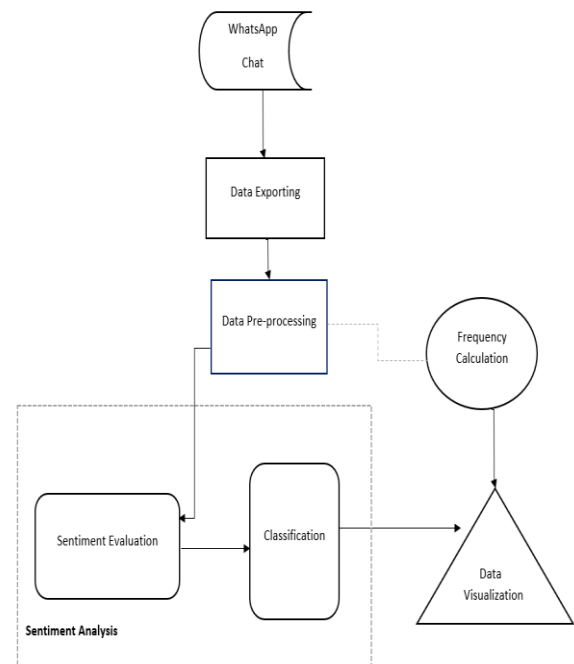
Figure 1: Shows whole process of the proposed worked starting from data collection to visualization

Methodologies :

a)Highest Emotion selection:

In this process the maximum number of repeated emotion will be considered as the overall emotion and no other factors are considered .Here total number for each emotion will be calculated for visualization purpose so it will provide us with the highest

b)Priority based model:

In this model all the emotions are given a priority number and based on the frequency of each emotion and its priority number overall score is calculated and with that overall emotion is calculated.

In t/his paper we worked with 2 models .

Scoring Range

Range	Emotion
<-50	Frustrated or more fear

-50 to -10	Sad and fear
-10 to 0	Little sad and un happy
0 to 10	Happy and little surprised
10 to 50	Surprised or excited or super happy
>50	Super excited

Table-1 Shows the scoring range

Model 1-

- Angry-2(-ve)
- Sad-2(-ve)
- Fear -5(-ve)
- Happy-1
- Surprised-3
- Frustrated-5(-ve)
- Excited -1
- Neutral -1

Model 2-

- Angry-5(-ve)
- Sad-2(-ve)
- Fear -2(-ve)
- Happy-2
- Surprised-2
- Frustrated-3(-ve)
- Excited -5

In the same way many number of models can be created and tested.

Results and analysis :

Now when normal method based on highest emotion selection is used below are the results

This data is processed on the chat data that have been collected from a one to one chat and pre-processed and visualised
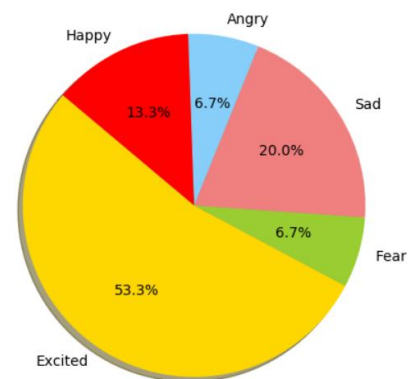
of the person A is on the particular chat and various emotions are displayed using various colours.

The below pie chart is the how the emotion

Figure-2 The above pie chart shows the variations of the emotions of the person
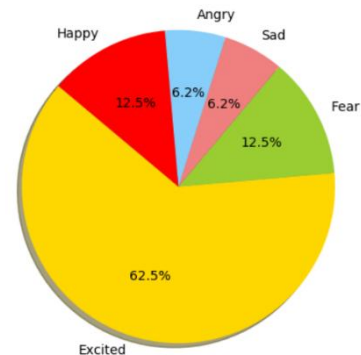
A

B

Figure-3 The above pie chart shows the variations of the emotions of the person

Person	Overall emotion	Messages per minute	Frequency
Person A	Excited	2.50	128
Person B	Excited	2.42	129

Table-2This table shows the metrics

Messages per minute is the average number of messages sent by a individual with in minute.

Person	Angry	Sad	Fear	Happy	Surprised	Frustrated	Excited
A	9	25	9	17	0	0	68
B	8	8	10	16	6	0	81
Total	17	33	25	33	0	0	149

Table-3This shows how the emotions are varying

Now using the second methodology we ran the two models with the different priorities and below are the results

Model-1

Person	Overall score	Overall Emotion
A	-28	Sad and Fear
B	3	Happy and surprised

Table-4 Shows the results from the model-1

Model-2

Person	Overall score	Overall Emotion
A	234	Extremely happy and Excited
B	353	Extremely happy and Excited

Table-5 Shows the results from the model-2

From the both the models we can say model 1 worked well based on the human instinct and also it is relatable to first methodology.

Future Work:

We are working on a model for estimating the accuracy based on the user facial emotions by using the front facing camera and real time analysis of the emotion of the person with the face and also with the text messages he has been typing and also the message that has been erased after typing the text area will be recorded and processed for better understanding of the user style of writing and also understanding the user mercurial behaviour. Also this whole model can be used on group chats as well where complex analysis of chats has to be made. Also this whole model will be implemented in the employees chat group to identify the particular person's activeness and involvement and frequency of texting. Also we are eliminating emojis while processing ,emojis in the 21st century means a lot and tells a lot about the user emotion.

Conclusion:

In conclusion we can argue for days stating this is a breach of privacy to analyse a person chat and process the emotions but when we see the brighter side of the project ,we can help people develop their character by helping them suggesting proper vocabulary to guide them in a proper path to reach a sustainable position with the other end person in terms of relations, business agreements and many more .Also this requires huge amount of data to master the art emotion is relative to each person and varies so the system has to be trained in a particular way to keep track of the context and past history while determining the sentiment. This paper proposes two new methods namely highest emotion selection and priority based model. Both has their own merits and demerits. But when results are analysed priority based model showed better promising results.

References :

1. Joshi, Sunil. (2019). Sentiment Analysis on WhatsApp Group Chat

Using R. 10.1007/978-981-13-6347-4_5.

2. Kaur, R. (2019). Insight to Emotional tones in WhatsApp Through Sentiment Analysis. IJRAR.

3. Winarko, E., &Cherid, A. (2017, November). Recognizing the sarcastic statement on WhatsApp Group with Indonesian language text.In 2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP) (pp. 1-6).IEEE.

4. Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. new media & society, 20(5), 1813-1831.

5. Preotiuc-Pietro, D., et al.: Trendminer: an architecture for real time analysis of social media text. In: Proceedings of the Workshop on Real-Time Analysis and Mining of Social Streams (2012)

6. Gupta, P., & Nene, M. J. (2017, March). Analysis of Text Messages in Social Media to Investigate CyberPsycho Attack.In International Conference on Information and Communication Technology for Intelligent Systems (pp. 581-587).Springer, Cham.

7. Medhat, W., Hassan, A., &Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

8. Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6), 282-292.

9. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.

10. Mäntylä, M. V., Graziotin, D., &Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27, 16-32.

11. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

12. Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338.

13. Rambocas, M., & Gama, J. (2013). Marketing research: The role of sentiment analysis (No. 489). Universidade do Porto, Faculdade de Economia do Porto.

14. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

15. Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior.Journal of management information systems, 29(4), 217-248.

16. Toubiana, M., &Zietsma, C. (2017). The message is on the wall? Emotions, social media and

the dynamics of institutional complexity. Academy of Management Journal, 60(3), 922-953.

17. Tauch, C., &Kanjo, E. (2016, September). The roles of emojis in mobile phone notifications. In Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing: Adjunct (pp. 1560-1565).

18. Resende, G., Melo, P., CS Reis, J., Vasconcelos, M., Almeida, J. M., &Benevenuto, F. (2019, June). Analyzing textual (mis) information shared in WhatsApp groups. In Proceedings of the 10th ACM Conference on Web Science (pp. 225-234).

19. Kalra, S., Johari, R., Dahiya, S., & Yadav, P. (2018). WAPiS: WhatsApp Pattern Identification Algorithm Indicating Social Connection. In Advanced Computational and Communication Paradigms (pp. 595-603). Springer, Singapore.

20. Mbunge, E., Vheremu, F., &Kajiva, K. A Tool to Predict the Possibility of Social Unrest Using Sentiments Analysis-Case of Zimbabwe Politics 2017-2018.

21. Reddy, C. B. S., Chaudhary, S., & Kandasamy, S. K. Spam, a Digital Pollution and Ways to Eradicate It.

22. Karamibekr, M., &Ghorbani, A. A. (2012, December). Sentiment analysis of social issues.In 2012 International Conference on Social Informatics (pp. 215-221).IEEE.

23. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.

24. El-Din, D. M. (2016). Enhancement bag-of-words model for solving the challenges of sentiment analysis. International Journal of Advanced Computer Science and Applications, 7(1).