Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Text Classification based Behavioural Analysis of WhatsApp Chats

Sonika Dahiya, Astha Mohta, Atishay Jain
Computer Engineering Department
Delhi Technological University
Main Bawana Road, New Delhi, India
sonika.dahiya11@gmail.com, asthamohta@gmail.com, atishay9jain@gmail.com

Abstract— **WhatsApp is used by millions of users to express emotions and share feelings. The model is presented in this paper aims to perform sentimental and emotional analysis using textual messages and emojis used in WhatsApp chats. Code switching, which is quite prevalent over online conversations, is handled by the model by unifying and converting all the texts to a standard form. For each subject, multiple chats are taken; translated and using a neural network, each sentence and emoji is scored in a dimensional form. The composition of the emotions expressed by the subject (out of Happy, Sad, Bored, Fear, Anger and Excitement) are defined. The scores are added up for each subject. Throughout the analysis, the behavioral traits are extracted. It is determined that, if the subject likes to use emojis and if they use it as a replacement for words or as an add-on to express their emotions better. It is also observed that if the subject behaves differently on text according to the person in front of them with regard to these emotions and finally, if the subject is an introvert or extrovert.**

Keywords— WhatsApp; Emoji; Emotion Analysis; Sentiment Analysis; Code-Mixing; Interpersonal Communication, Behavior Analysis

I. INTRODUCTION

WhatsApp has secured its place as the most prevalent social media service in the world with two billion users [1]. WhatsApp connects users across 180 countries with one another and over 55 million messages are sent on WhatsApp every single day [2]. WhatsApp is an integral part of people's lives, not only to share information but also to share feelings through meaningful conversations.

A study was conducted in Chennai, India on students to investigate the prominence of WhatsApp among youth. The results indicated that subjects stay online for sixteen hours in a day and spend almost eight hours in a day on WhatsApp [3]. Thus, it is concluded that WhatsApp chats can help depict a person's emotions and capture their sentiments and thought process.

Sentiment analysis can be described as techniques, methods, and tools for detecting as well as extracting subjective information, such as attitude and opinions, from language. Until recently, sentiment analysis has been about opinion polarity, i.e., categorization of text as positive, negative or neutral [4]. However, these three conventional categories are not adequate to recognize the connotation of the underlying tone of a sentence [5].

Hence, a novel approach using both sentiment analysis and emotion analysis are provided. Sentiment analysis categorizes emotions as positive, negative and neutral. Emotion analysis detects the type of feelings and identifies the temperament of the user as a combination of emotions, happy, sad, angry, fear, excitement or boredom. Data for the study was collected through crowd sourcing. People of different professional backgrounds and age groups were approached and with their consent asked to share their chats for the purpose of this study. The data is analyzed using in order to derive various features to guide the behavioral analysis.

In this paper, the previous work on the Emotion Classification, WhatsApp Chat analysis as well as texting psychology is introduced in Section 2. Then the Data Set along with the proposed methodology for collection and key insights and pre-processing of the data in Section 3. The proposed algorithm for calculation scores for each individual subject is discussed in section 4. Section 5 discusses external tools (Parallel Dot) used for multi-classification of emotions and Validity check performed on the Library. Finally the results are presented in section 6 followed by the conclusion in section 7.

II. RELATED WORK

WhatsApp originated as a free substitute to regular SMS. Today it can be used for sending variety of media and, making voice and video calls [2].

As per the study conducted by the authors of "Survey Analysis on the usage and Impact of WhatsApp Messenger" on a set of WhatsApp users between the ages of 18 and 50, it was recorded that about 79% of the participants use WhatsApp for at least fifteen to sixty minutes daily [6]. In another study conducted by the authors of "Impact of WhatsApp on youth: A Sociological Study" hundred random WhatsApp users in the city of Agra were selected between the age of 18 and 30. Their study shows that 63% users used WhatsApp at a frequency of 50 times a day and 21% users used it at frequency of 20 times a day while the rest 16% used it at a frequency of 100 times a day. It is concluded from this study that among the youth, WhatsApp has a huge impact on how they communicate. These users stated that about 40% of their WhatsApp list was close friends and work colleagues and acquaintances consist of 25% and 20% respectively [7].

A study conducted on students Abu Dhabi, reveals that 85% female students and 70% male students use emojis to replace facial expression on text. These students used WhatsApp every day for an average of 1 to 7 hours. Hence, emojis is an important for communication over WhatsApp [8]. In another study where 3.88 million users from various countries were studied, it was revealed that 7.01% of their 6.06 billion messages contained at least one emoji. This study reveals how popular usage of emojis actually is among users [9].

In the paper "Facebook Sentiment: Reactions and Emojis", it is presented that emojis and the linguistic text modify each other's meaning [10]. The emojis interact with the linguistic text in many ways; an emoji can replace a word, repeat a phrase, independently express the emotion or attitude of the speaker, emphasize an emotion present in a text or can be used to be more polite. The paper concludes by suggesting that emojis can be used to detect user's sentiment if the context where their meaning is modified is also taken into account.

For the study purpose, each sentence is evaluated and emoji as a degree of six emotions. Russell's Circumplex approach suggests that experiences can be represented using a multi-dimensional model as a linear combination of degree of valence (pleasure-displeasure) and arousal (activation-deactivation) [11]. As an example, Excitement is considered as an emotional state with high pleasure, high arousal. On the other hand, Joy is considered as an emotional state with positive valence, i.e. pleasure and moderate arousal [12].

To cover the scale, six emotions that cover all degrees of arousal and valence. Anger and fear lie in the quadrant of activation and displeasure, sadness and boredom in deactivation and displeasure, happiness and excitement in activation and pleasure. Deactivation and pleasure is satisfaction and leads to a neutral state of mind. In our first step of the algorithm, the neutrality has been checked [13-14]. Figure 1 shows these emotions on a two dimensional space.

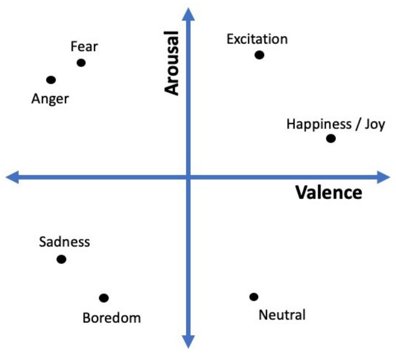


Fig. 1. Categorization of emotion based on Russell's Circumplex model

Aravind K. Joshi describes code-switching or code mixing as "Speakers of certain bilingual communities systematically produce utterances in which they switch from one language to another in the course of an utterance. Production and comprehension of utterances with intra sentential code-switching is part of the linguistic competence of the speakers and hearers of these communities." [15]. "Code Mixing: A Challenge for Language Identification in the Language of Social Media" discusses how despite most social media content being in English, this still only makes up for half of it as users switch between language mid-way. [16]. Hence some form of translation or model which is designed for tagging of multiple languages is required.

The emotions can be classified as Categorical or Dimensional. In categorical, emotions are divided into groups and assigned into a discrete category, such as positive or negative one or as Ekman suggested happiness, sadness, fear, disgust, anger, & surprise. Limitation of categorical emotion classification is that it prohibits reflection of demeanor and degree of an emotion and translation may lead to reduced informational value due to language specific meanings. In comparison, for multi-dimensional categorization does not face this issue as it is subject of assignment to many variables and not a single category [17]. The model prescribed in this paper is an extended model built on top of the above two models.

Previous works for analyzing WhatsApp chats use a bag-of-words approach where frequency use of words is used [18]. The bag-of-words can be used to train a model to predict the emotion category. Another strand of bag-of-words includes using the "syuzhet" package in R language [19]. The final sentiment score is calculated by finding all the lexicon word in each sentence in the input vector and algebraically adding those [20]. Another approach uses AFINN lexicon in which each word is assigned a score between -5 and 5. Along with this common words used by the user, frequency of messages on each day, emoji classification and lexical diversity assist the study [21]. The limitations of the discussed words are that the categorical classification as proposed by Ekman completely disregards the possibility of multiple emotions being present in a single sentence. Furthermore, none of them take into consideration the challenge of code switching.

III. Data Collection

A. Data Collection and User Consent

For the purpose of the research, eighteen subjects from different backgrounds and age groups were asked to share their WhatsApp conversations with 5-10 people each. As the proposed work crowdsources the data, confidentiality and privacy of the data needs to be respected. Firstly, all the WhatsApp Chats have been collected with signed consents from the users and none have been made public. Furthermore, in scenarios where the scalability of the experiment has to be increased, other methods to increase the protection of privacy of the conversations such as data perturbation, data anonymization, and cryptographic techniques may also be used. The "Export Chat" feature available on WhatsApp has been used to collect the conversations via Email, WhatsApp, Bluetooth, etc. [22-23]. All the subjects were asked to share their conversations without media, since the model requires only textual data.

B. Analysis

A database is formed consisting of around 180 conversations of 18 subjects. General demographic information including name, age, gender, profession was also collected. Analysis has been done on the basis of professional fields and age groups to project the diversity of the subjects. This analysis is shown in table 1 and 2.

TABLE I. NUMBER OF SUBJECTS BELONGING TO DIFFERENT AGE GROUPS

Age Groups	Number of subjects
15- 25	8
25- 35	2
35- 45	2
45- 55	5
>55	1

TABLE II. NUMBER OF SUBJECTS BELONGING TO DIFFERENT PROFESSIONAL BACKGROUNDS

Field of Profession/ Education	Number of Subjects
Engineering	5
Medicine	4
Humanities	4
Commerce	5

It has been studied that emojis can often act as a replacement for words when words are not sufficient to express our emotions [24]. Also emojis express how comfortable the users are in a conversation and emoji count is inversely proportional to how professional the conversation between the two parties is [25]. Another study on how often introverts and extroverts tweet shows that introverts tweet 14.4% more often in a fixed duration of period [26]. Along with this, texting is a preferred mode of communication over phone calls for introverts [27]. Hence, the two subjects are analyzed on the basis of their emoji count, word count, sentence count and the number of emojis used per sentence.

Consider two Subjects from the dataset and compare their chats on the bases of emoji count, word count, sentence count and the number of emojis used per sentence. Table 3 gives a comparison based on words, sentences, emojis (along with categories) and emoji to sentence ratio.

TABLE III. ANALYSIS OF TWO SUBJECTS ON EMOJI, WORD, SENTENCES AND EMOJI:SENTENCES COUNT

Counts	Subject 1	Subject 2
Emoji Count	183	855
Word Count	3472	17221
Sentence Count	938	3440
Emoji: Sentence Count	0.195	0.259

It can be seen clearly that Subject 2 has a higher emoji count per sentence as compared to subject 2. Hence according to the studies explored above it can be said that subject 1 mostly communicates professionally and is less expressive online whereas subject 2 with a score of 855 generally communicates unprofessionally and is quite expressive. It is also said that by looking at the sentence count that texting is a preferred mode of communication for Subject 2. It is generally seen that texting is more popular among introverts as compared to extroverts.

C. Pre-Processing

1) Emojis: The entire set of emoji codes is defined by the Unicode consortium. The emoji list prescribed by Unicode along with the keywords associated with every emoji is used to classify them into 6 categories. WhatsApp allows different skin colors for certain emojis. Hence Regex library in combination with the Unicode Library can be used to extract these emojis [28]. The classification is shown in figure 2.

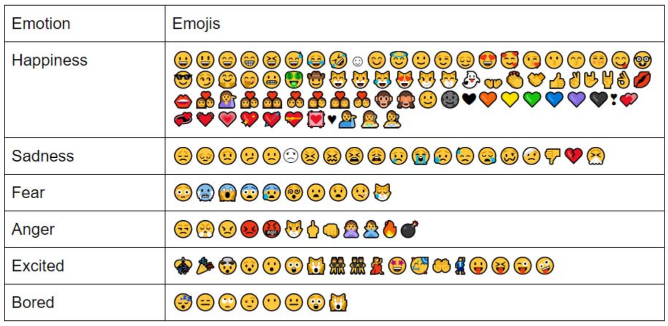


Fig. 2. Emojis categorized by emotions

2) Steps in Pre-Processing:

a) Regex library is used in conjunction with the emoji library to find emoji patterns in the text and separate the text from emojis.

b) Following this, a random fixed number of sentences are selected per chat from the texts and GoogleTrans (python library that implemented Google Translate API) is used to translate them in to English language [29].

IV. PROPOSED ALGORITHM

A. Algorithm

In the first step WhatsApp chats are collected from various users. These chats include messages from both sender and receiver with time stamps. Since it is only required to analyze sender chats, the chats of intended subject are separated and store them separately in the second step.

Next, each sentence of the timestamp are stripped and characters are checked to segregate sentences from emojis. Both emojis and sentences are stored and processed separately. Emojis are now classified into one of the 6 Emotions Classes: Anger, Happiness, Sadness, Fear, Excited and Bored.

To process sentences, first translate each sentence using Google Trans as sentences are in a combination of English and Hindi. After this each of the translated sentences is passed through a Neutrality Classifier. Here the sentence maybe classified as Positive, Negative or Neutral. To reduce the effect of Neutral on our scores, calculate the Weightage Factor as {1.0-N}. N is the Neutrality Score over here. Hence, sentences which are less Neutral will affect the score more than a less Neutral sentence. Now classify each sentence into 6 Emotion Classes. Each Emotion receives a value out of 1.0, which is the probability of it expressing that emotion. Hence, multiply this value to the Weightage Factor (WF).

Sum of all emojis Scores and Sentence score are finally added together, to receive the final score. These values are plotted to examine and compare a person's emotional scale. The algorithm is depicted in Figure 3.

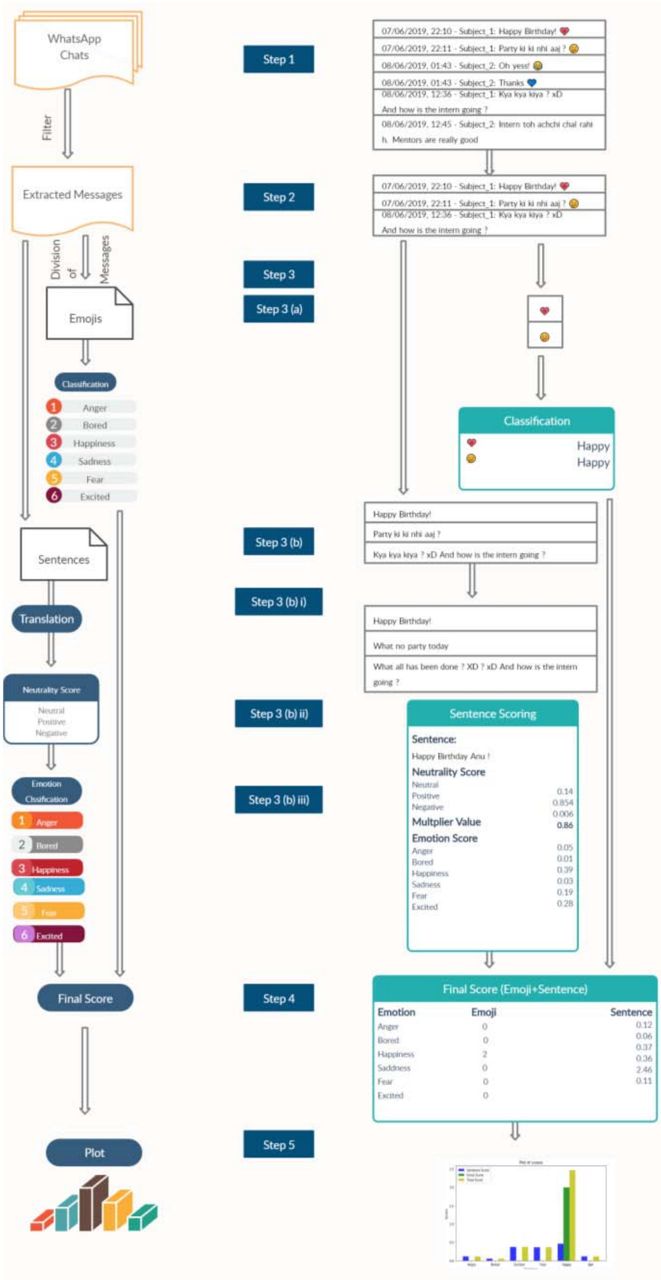


Fig. 3. Algorithm with Sample Chat

B. *Scoring*

1) Emoji Scoring

Emotions are portrayed by emojis unambiguously. Thus the emojis is a perfect indicator of the emotion and is topic as well as domain independent [30]. Quite often though, a single text message may relentlessly use emojis. After experimenting with different values, a threshold value of 3 was chosen. Usage of the same emoji > 3 times consecutively implies a loss of weightage. This fact is incorporated by dividing the emoji value count by 3 and adding the ceiling value in case emojis of same category are used > 3 times. Initial Emoji Count for each emotion is taken to be equal to zero. As the emojis are encountered, the respective counts of categories are increased in accordance with the classification table discussed in section 3 C.

Let, Emoji Count$_i$ = EC_i

for i=[Anger, Bored, Happiness, Sadness, Fear, Excited]

then,

$$f(EC_i) = \begin{cases} \lceil EC_i/3 \rceil, & EC_i > 3 \\ EC_i, & otherwise \end{cases} \quad (1)$$

2) Sentence Scoring

Sentence Scoring is done in two parts: Calculating Neutrality and Calculating Emotion Strength. This is done because some sentences show higher emotional intensity than others. Since one sentence is considerably more neutral than other, it is important to weigh this in. Hence, a weighting factor is used to add a weight to each Emotion. Weighting Factor (WF) and sentence score is calculated as:

$$Weighing\ Factor\ (WF) = 1.0 - Neutrality \quad (2)$$

$$\begin{pmatrix} Total\ Sentence\ Anger\ Score = WF * Sentence\ Anger\ Score \\ Total\ Sentence\ Excited\ Score = WF * Sentence\ Excited\ Score \\ Total\ Sentence\ Sad\ Score = WF * Sentence\ Sad\ Score \\ Total\ Sentence\ Happy\ Score = WF * Sentence\ Happy\ Score \\ Total\ Sentence\ Bored\ Score = WF * Sentence\ Bored\ Score \\ Total\ Sentence\ Fear\ Score = WF * Sentence\ Fear\ Score \end{pmatrix} \quad (3)$$

3) Total Score

Total Score is calculated as:

$$\begin{pmatrix} Total\ Anger\ Score = \sum Sentence\ Anger\ Score + \sum Emoji\ Anger\ Score \\ Total\ Excited\ Score = \sum Sentence\ Excited\ Score + \sum Emoji\ Excited\ Score \\ Total\ Sad\ Score = \sum Sentence\ Sad\ Score + \sum Emoji\ Sad\ Score \\ Total\ Happy\ Score = \sum Sentence\ Happy\ Score + \sum Emoji\ Happy\ Score \\ Total\ Bored\ Score = \sum Sentence\ Bored\ Score + \sum Emoji\ Bored\ Score \\ Total\ Fear\ Score = \sum Sentence\ Fear\ Score + \sum Emoji\ Fear\ Score \end{pmatrix} \quad (4)$$

V. EXPERIMENTAL SETUP

A. *Parallel Dots*

Parallel Dots is an Artificial Intelligence Platform created Using advanced Deep Learning and Machine Learning techniques. The classifier uses Convolutional Neural Networks on a self-classified dataset. Valence – Arousal scores to capture emotion are used to label most Datasets [31].

Sentiment Score API returns confidence score for each of the tags, positive, negative and neutral [32]. Emotion Score API returns confidence score for each of the tags: Happy, Sad, Angry, Excited, Sarcasm or Fear [33].

Parallel Dots runs on a Deep Learning algorithm. The emotion associated with the text is classified based on features which are extracted from the textual data. The classifier is trained on a labeled dataset using Convolutional Neural Networks (CNN).

B. Validity

In order to validate the prescribed model, a data set pre-classed into the 6 emotions was required. The Data Set is acquired from Kaggle [34]. The data set contains 29939 unique statements classified into 13 emotions (Happy, Sad, Angry, Bored, Excited, Fear, Empty, Neutral, Fun, Love, Hate, Surprise and Relief). The data set is filtered to contain only the sentences belonging to the 6 emotions classified by the model prescribed in the paper.

Accuracy is calculated on the basis of total number of sentences correctly classified by the classifier.

$$Accuracy = \frac{total\ number\ of\ sentences\ correctly\ classified}{total\ number\ of\ sentences} \quad (5)$$

To validate the model, if the emotions corresponding to the top score matched the classified emotion in the data set, then that statement is categorized as correctly classified otherwise not. The above algorithm is applied on the data set in batches and average accuracy is found out to be 72.9%.

For comparison an SVM (Support Vector Machines) is used as described in previous works. A text processor function is created which removed the punctuation and stop words. A pipeline is created consisting of the following steps. CountVectorizer with the above text processor function passed as analyzer is applied to convert strings to token integer counts. TF-IDF Transformer is applied to convert integer counts to weighted TF-IDF scores. SVM classifier is applied. The accuracy obtained by using the above model is 33.2%.

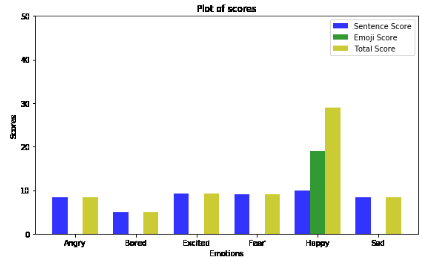
VI. RESULTS

A. Comparative Study of Four users

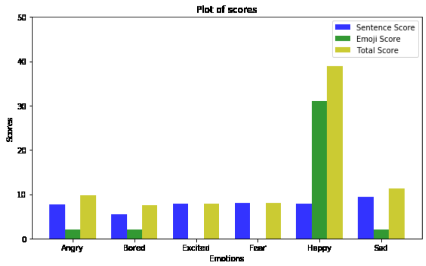
The algorithm proposed in section 4 A was run on the complete data set consisting of 5 – 10 chats each of 18 different subjects (Data set is extensively described in section III). In Figure 4, 5, 6 and 7 present the results as well as a comparative study of 4 of the 18 subjects. The graph plots the Total Sentence Score, The Total Emoji Score and the Total Score (Sentence Score + Emoji Score) for each chat per subject. For each subject, chats with 4 people are taken (hence 4 graphs per subject) and the graphs have been normalized in the range of 0 – 50 for ease in comparison. Equal number of sentences were chosen so that their total scores could be compared.

Following this the hypothesis on the person's behavior is compared with the opinion have received from our primary and secondary sources. The primary source is the subject themselves and the secondary source is the user the subject interacted with in the chats. According to Simine Vazire, Ph.D., Washington University "While individuals may be more accurate at assessing their own neurotic traits, such as anxiety, it seems friends, and even strangers, are often better barometers of traits such as intelligence, creativity and extroversion" [35]. Graphs were plotted as depicted in figure 4, 5, 6 and 7 and analysis was performed based on this.

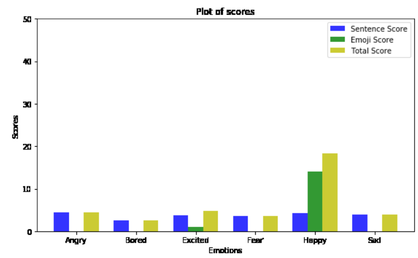
a)



b)



c)



d)

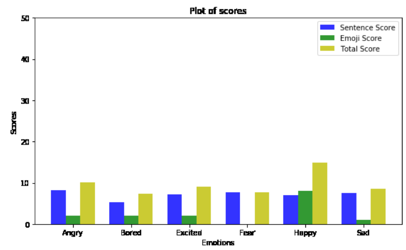
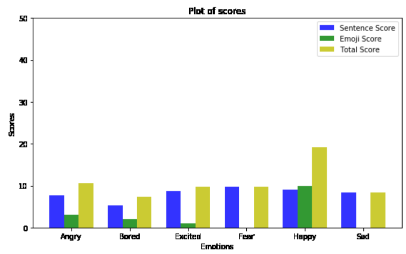
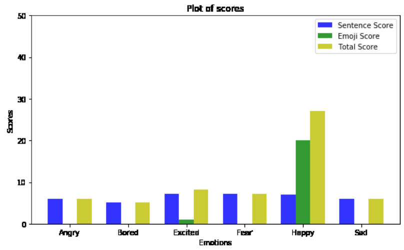


Fig. 4. Subject A Graphs for 4 chats

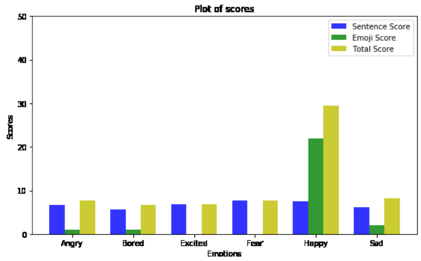
a)



b)



c)



d)

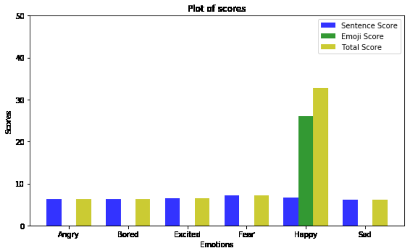
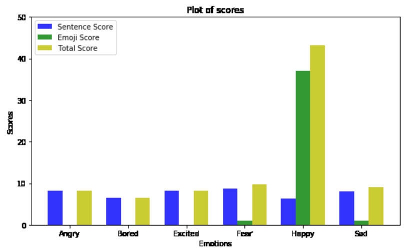
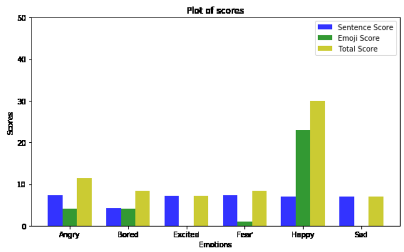


Fig. 5. Subject B Graphs for 4 chats

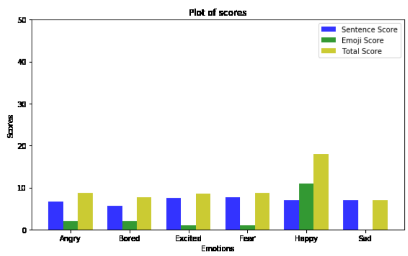
a)



b)



c)



d)

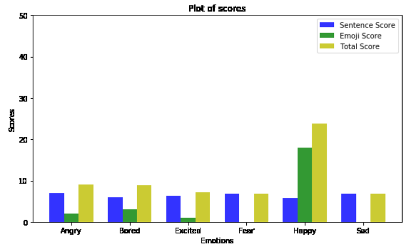
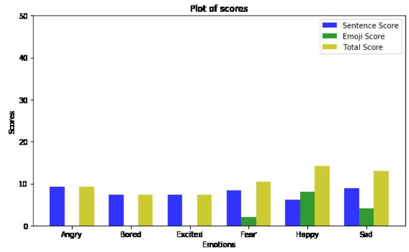
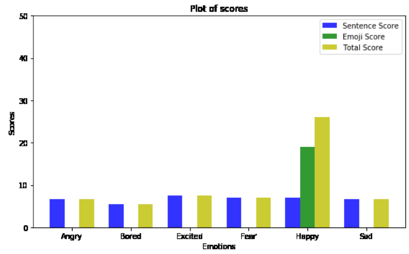


Fig. 6. Subject C Graphs for 4 chats

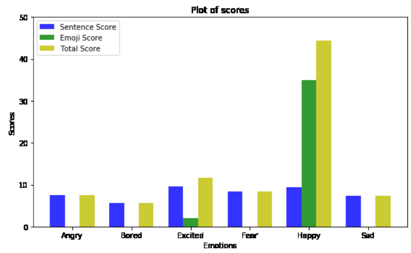
a)



b)



c)



d)

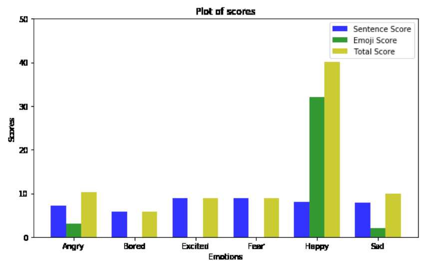


Fig. 7. Subject D Graphs for 4 chats

For each subject, the following questions were asked to the subject themselves (Primary Source Opinion) and to the people those they text (Secondary Source Opinion):

1. Is the subject introverted or extroverted?

2. Do they like to use emojis? Why do they use the emojis, as a replacement of words or and add on?

3. Which emotions do they express more? Answer with the top three among the following along with how strongly

they express it. Emotions are Happy, Sad, Bored, Fear, Anger and Excitement.

4. How differently does the subject behave on text according to the person in front of them with regard to these emotions?

Table 4 shows the comparison drawn for each of the four test subjects on the result according to our hypothesis, primary source opinion and secondary source opinion.

TABLE IV. COMPARATIVE BEHAVIOR ANALYSIS OF 4 SUBJECTS

Subject	Hypothesis on the basis of Results	Primary Source Opinion	Secondary Source Opinion
A	Subject A is extroverted. Subject A shows less emotions on text. They usually only use Happy emojis. This shows that they use emojis as an add-on and not to replace words. The subject behaves differently according to the person they are interacting with. The subject is usually Happy or Excited but also portrays Anger.	Subject A feels they are extroverted. They use emojis as an add-on because they are emojis are fun. Subject A feels they are usually Happy and Excited and sometimes Bored. They feel they portray their emotions moderately. They feel they behave similarly with everyone.	Secondary sources feel Subject A is extroverted. They use emojis extensively but as an add-on for fun. Subject A is usually Happy or excited and sometimes bored. They act similar with whoever they interact with.
B	Subject B is introverted. Subject B's range of emotion does not vary highly with the person they are interacting with. They prefer not to show too many emotions on text. With some people they show only positive emotion while with others they express all emotions equally. Subject B shows high values of Happiness and Fear. Also, subject B uses emojis only as an add-on and not to replace words.	Subject B feels that they are introverted. They feel they show Happiness and Fear along with Sadness sometimes. They prefer to behave similarly with everyone they interact with and do are not highly expressive of their emotions on text. They use emojis as an add-on and not replacement of words.	Subject B is introverted. They usually show Happiness, Excitement and Fear. They may show sadness at times but prefer to hide emotions on text. Hence, they behave similarly with whoever they are texting. They use emojis but only as an add-on and not as a replacement for words.
C	Subject C is extroverted. Subject C likes to replace words with texts, as they use all types of Emojis to express their emotions. This can also be derived from the fact that even though subject C's happiness score is low; they use emojis to express Happiness. They show all emotions except boredom. Subject C behaves similarly with all people.	Subject C feels they are extroverted. They use emojis to replace words as they are highly expressive of their emotions and emojis help them explain these better. They feel they show Happiness, Sadness, Excitement and Anger. They feel they are lively, happy and cheerful and behave similarly with all people.	Subject C is extroverted. Emojis help them express their emotions better and are used as a replacement for words. They are usually Happy, Excited, Sad or Angry. They behave similarly to everyone they talk to.
D	Subject D is extroverted. Subject D is expressive in their messages. They use show Happiness, Excitement and Fear. Their interaction style differs according to the person they interact with. They use emojis to replace words and emojis help them express their emotions better.	Subject D feels they are extroverted. Their emotions and quirkiness depend solely on the person they interact with. They use emojis when they are usually excited. They help them express their emotions better. They are usually Happy, Excited or Sad.	Subject D is extroverted. If they are close to someone, they would not hesitate to express their emotions. Hence, they act differently with different people and their emotions vary according to the person and their mood. Subject is very expressive in emojis and is usually Excited, Sad or Happy.

B. Accuracy

The algorithm proposed in Section 4 A is applied to all 18 subjects on 10 chats each. For each chat, number of words, sentences, emojis and their ratio is computed. Following this the algorithm is run on a fixed number of sentences (1000) per chat. Graphs are plotted for the same and questions mentioned in the last section are answered for each subject. Table 5, 6, 7 and 8 show accuracy of our analysis for each question by matching our hypothesis with primary and secondary opinion.

TABLE V. ACCURACY FOR QUESTION 1

Question 1. Is the subject introverted or extroverted?		
Does Hypothesis match with primary and secondary opinion	**Count**	**Accuracy%**
With both primary and secondary opinion	13	72.2%
With only primary opinion	1	5.5%
With only secondary opinion	3	16.6%
With neither primary nor secondary opinion	1	5.5%

TABLE VI. ACCURACY FOR QUESTION 2

Question 2. Do they like to use emojis? Why do they use the emojis, as a replacement of words or and add on?		
Does Hypothesis match with primary and secondary opinion	**Count**	**Accuracy%**
With both primary and secondary opinion	15	83.3%
With only primary opinion	0	0.0%
With only secondary opinion	0	0.0%
With neither primary nor secondary opinion	3	16.6%

TABLE VII. ACCURACY FOR QUESTION 3

Question 3. Which emotions do they express more? Answer with the top three among the following along with how strongly they express it. Emotions are Happy, Sad, Bored, Fear, Anger and Excitement.		
Does Hypothesis match with primary and secondary opinion	**Count**	**Accuracy%**
With both primary and secondary opinion	9	50.0%
With only primary opinion	1	5.5%
With only secondary opinion	2	11.1%
With neither primary nor secondary opinion	6	33.3%

TABLE VIII. ACCURACY FOR QUESTION 4

Question 4. How differently does the subject behave on text according to the person in front of them with regard to these emotions?

Does Hypothesis match with primary and secondary opinion	Count	Accuracy%
With both primary and secondary opinion	8	44.4%
With only primary opinion	3	16.6%
With only secondary opinion	4	22.2%
With neither primary nor secondary opinion	3	16.6%

VII. CONCLUSION AND FUTURE WORK

A. Conclusion

The chat analyzer is a gateway to the unexplored field of WhatsApp Chats. It allows to analyze bilingual user. Most work done in emotion analysis focuses on classifying the user's emotions as Positive, Negative or Neutral. The Chat Analyzer goes a step ahead and classifies these emotions into six different emotions and uses their Neutrality to weigh them. It also classifies different emojis as emojis are a popular way of expressing emotions. Our Analyzer gave 72.9% accuracy against a set of pre-classified data.

A rise in hate speech, cyber-bulling, heckling and increased impudence on social media has been observed. The Chat Analyzer can be used as a tool to give a user an insight on their online behavior as they communicate with their peers. It allows a user to keep a check on their emotions by analyzing them.

The model described in this paper is successfully applicable to WhatsApp Chats, classifies the texts into one of 6 emotions while taking into consideration the emojis used by the person and therefore stands apart.

REFERENCES

[1] WhatsApp (2020) About WhatsApp. WhatsApp. Available: https://www.whatsapp.com/about/

[2] J. Andjelic (2019) WhatsApp Statistics: Revenue, Usage, and History. Fortunly. Available: https://fortunly.com/statistics/whatsapp-statistics#gref

[3] Ka, Jisha & Jebakumar, Dr. (2014). Whatsapp: A Trend Setter in Mobile Communication among Chennai Youth. IOSR Journal of Humanities and Social Science. 19. 01-06. 10.9790/0837-19970106.

[4] Sayali Meshram, Ms. Nisha Balani, Ms. Parul Jha (2019). Sentiment Analysis of Statement. IOSR Journal of Engineering (IOSRJEN). Vol. 09, Issue 5 (May. 2019), ||S (XI) || PP 46-49

[5] Yaakub, Mohd Ridzwan & Abu Latiffi, Muhammad Iqbal & Safra, Liyana. (2019). A Review on Sentiment Analysis Techniques and Applications. IOP Conference Series: Materials Science and Engineering. 551. 012070. 10.1088/1757-899X/551/1/012070.

[6] Kumar, Naveen & Sharma, Sudhahsh. (2017). Survey Analysis on the usage and Impact of Whatsapp Messenger. Global Journal of Enterprise Information System. 8. 52. 10.18311/gjeis/2016/15741.

[7] Bhatt, Anshu & Arshad, Mohd. (2016). Impact of WhatsApp on youth: A Sociological Study. IRA-International Journal of Management & Social Sciences (ISSN 2455-2267). 4. 376. 10.21013/jmss.v4.n2.p7.

[8] Kootbodien, Ammaarah & Prasad, Nunna & Ali, Muhamad. (2018). Trends and Impact of WhatsApp as a Mode of Communication among Abu Dhabi Students. Media Watch. 9. 10.15655/mw/2018/v9i2/49380.

[9] Lu, Xuan & Ai, Wei & Liu, Xuanzhe & Li, Qian & Wang, Ning & Huang, Gang & Mei, Qiaozhu. (2016). Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. 770-780.

[10] Tian, Ye & Galery, Thiago & Dulcinati, Giulio & Molimpakis, Emilia & Sun, Chao. (2017). Facebook sentiment: Reactions and Emojis. 11-16. 10.18653/v1/W17-1102.

[11] Russell, James. (1980). A Circumplex Model of Affect. Journal of Personality and Social Psychology. 39. 1161-1178. 10.1037/h0077714.

[12] Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol. 2005;17(3):715–734. doi:10.1017/S0954579405050340

[13] Acampora, Giovanni & Loia, Vincenzo & Vitiello, Autilia. (2011). A cognitive multi-agent system for emotion-aware ambient intelligence. IEEE SSCI 2011 - Symposium Series on Computational Intelligence - IA 2011: 2011 IEEE Symposium on Intelligent Agents. 10.1109/IA.2011.5953606.

[14] Zhao G, Zhang Y, Ge Y. Frontal EEG Asymmetry and Middle Line Power Difference in Discrete Emotions. Front Behav Neurosci. 2018;12:225. Published 2018 Nov 1. doi:10.3389/fnbeh.2018.00225

[15] Aravind K. Joshi. 1982. Processing of sentences with intra-sentential code-switching. In Proceedings of the 9th conference on Computational linguistics - Volume 1 (COLING '82). Academia Praha, CZE, 145–150. DOI:https://doi.org/10.3115/991813.991836.

[16] Barman, Utsab & Das, Amitava & Wagner, Joachim & Foster, Jennifer. (2014). Code Mixing: A Challenge for Language Identification in the Language of Social Media. 10.13140/2.1.3385.6967.

[17] Bruna, Ondřej & Avetisyan, Hakob & Holub, Jan. (2016). Emotion models for textual emotion classification. Journal of Physics: Conference Series. 772. 012063. 10.1088/1742-6596/772/1/012063.

[18] Freeman Goja (2019) Sentiment Analysis Of WhatsApp Chat. Medium. Available: https://medium.com/analytics-vidhya/sentiment-analysis-of-whatsapp-chat-a906105917f4

[19] Joshi, Sunil. (2019). Sentiment Analysis on WhatsApp Group Chat Using R. 10.1007/978-981-13-6347-4_5.

[20] Naldi, M. (2019). A review of sentiment computation methods with R packages. ArXiv, abs/1901.08319.

[21] Keagan Stokoe (2020) Text and Sentiment Analysis of WhatsApp Messages. Available: https://levelup.gitconnected.com/text-and-sentiment-analysis-of-whatsapp-messages-1eebc983a58

[22] WhatsApp (2020) How to save your chat history. WhatsApp. Available: https://faq.whatsapp.com/en/android/23756533/

[23] WhatsApp (2020) How to save your chat history. WhatsApp. Available: https://faq.whatsapp.com/en/iphone/20888066/

[24] Alizah K. Lowell (2016) Why Do We Use Emojis?. Psychology Today https://www.psychologytoday.com/us/blog/contemporary-psychoanalysis-in-action/201605/why-do-we-use-emojis

[25] Haji, Hadi & Nawxosh, Salam. (2019). The Use of Emoticons among University Students: A Pragmatic Study. Zanco Journal of Humanity Sciences. 23. 10.21271/zjhs.23.1.19.

[26] Zhou, Z., Xu, K. & Zhao, J. Extroverts tweet differently from introverts in Weibo. EPJ Data Sci. 7, 18 (2018). https://doi.org/10.1140/epjds/s13688-018-0146-8

[27] Brandi Neal (2017) 7 Differences In The Way Introverts And Extroverts Communicate. Bustle. https://www.bustle.com/p/7-differences-in-the-way-introverts-extroverts-communicate-60756

[28] Python Software Foundation (2020) Regular expression operations. Python Software Foundation. https://docs.python.org/2/library/re.html

[29] SuHun Han (2018). googletrans 2.4.0. PyPi. Available: https://pypi.org/project/googletrans/

[30] Novak, P.K.; Smailovi´c, J.; Sluban, B.; Mozetiˇc, I. Sentiment of Emojis. PLoS ONE 2015, 10, e144296

[31] ParallelDots (2019) APIs Documentation. ParallelDots. Available: https://www.paralleldots.com/docs

[32] ParallelDots (2019) Text Analysis APIs. ParallelDots. Available: https://www.paralleldots.com/text-analysis-apis#sentiment

[33] ParallelDots (2019) Text Analysis APIs. ParallelDots. Available: https://www.paralleldots.com/text-analysis-apis#emotion

[34] Kaggle InClass competition (2017). Sentiment Analysis: Emotion in Text. Kaggle. Available: https://www.kaggle.com/c/sa-emotions/data

[35] Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of Personality and Social Psychology, 98(2), 281–