Emotional Conversational Analytics – An experimental case study

Dr. Jay Kiruthika
Kingston University
j.kiruthika@kingston.ac.uk

***Abstract*— Users are shifting to voice over message-based communications rather than typing. There exists a gap in evaluating emotional aspects of the phrases uttered. There are software analytics for sentimental analysis and other ways of measuring emotional aspects of the spoken phrases. But there is a lack of a system that evaluates secondary and primary emotions that can be used in various areas to detect human behavior. In order to facilitate this, there is a need to decipher the primary and secondary emotions of the user. This project addresses a novel idea of creating an emotional phrase map that identifies secondary and primary emotions, in addition to using machine language to train the system to identify the phrases and classify the emotions via a case study and the results analyzed.**

Keywords- Emotional Conversational analytics; Emotional Conversations; Emotional NLP;

I. Introduction

The global market for voice search devices grows every year as the devices are commercialized for budgeted households. Companies like Apple, Google, Xiaomi, Amazon and other manufacturers have flooded the market for this purpose. These devices have various functions and features that facilitate a user to voice search, find information, retrieve everyday planner, do video calling, communicate with each other etc. Users like to use voice communication as it is faster and need not be tied to a physical keyboard/device. The voice commands can be uttered in a physical area of coverage. Often sentiment scoring is used to determine primary emotions of the uttered words by the users. But secondary emotions are ignored. Using Robert Plutchik's[1] emotional wheel, primary and secondary emotions can be evaluated to learn more about emotional aspects of the voice chat or conversation. As future growth of these devices is expected to grow exponentially there is a need to elucidate emotions of the phrases uttered. This also can be used for monitoring and other purposes.

This project tends to use voice over chat conversations to provide intelligent conversational analytics solutions intended for various uses. For example, it can be used to identifying vulnerable users seeking help which otherwise is not identified by other means. It can also be used in detecting cyber threat/violence and abusive behavior. A smart notification system can be built on the resulting analytics that can be used to monitor the user's intent. This requires creating a new type of phrase map(s), named emotional phrase map(s) in this project that can help identify the primary and secondary emotions of the phrase uttered (Fig 1.1).

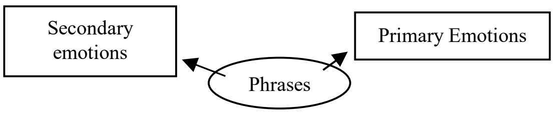

Fig 1.1 Learning Journey.

The NLP and NLU are currently used to decipher phrases, focus on understanding the language structure in addition to determine user's intent. Google, Amazon uses training data to achieve a higher success rate in determining user intent.

Plutchik[1] classifies primary & related secondary/ and the dyads of emotions that sit between two. This can be integrated to current NLP for deciphering human emotions.

Primary	Secondary	Sub sect of Secondary	Dyads of the subsect
Rage	Anger	Annoyance	Aggressiveness
Vigilance	Anticipation	Interest	Optimism
Ecstasy	Joy	Serenity	Love
Admiration	Trust	Acceptance	Submission
Terror	Fear	Apprehension	Awe
Amazement	Surprise	Distraction	Disapproval
Grief	Sadness	Pensiveness	Remorse
Loathing	Disgust	Boredom	Contempt
Rage	Anger	Annoyance	

Fig 1.2 Emotional classification with dyads

The current focus of an NLP[2] is to :

- Understand language semantics
- Understand User Intent
- Understand the phrases with variable synonyms etc via phrase maps.

II. Emotional Phrase Maps

The phrase maps are used to recognize and register the words that were being uttered by a user. This helps to understand the communication between a human and a

2473-3636/21/$31.00 ©2021 IEEE
DOI 10.1109/DCABES52998.2021.00028

device to provide meaningful feedback to the user. These phrase maps are predominantly used in creating voice bots and in CUI(Conversational User Interface) designs. A regular phrase map[4][6] contains insightful text analysis with machine language that extracts, analyzes and stores text. This can be used to train & customize applications and to classify, extract and detect sentiment with minimum effort using NLU[7] resulting in sentiment analysis, entity analysis, infographics, content classification etc.

Whilst creating an emotional conversational analytics platform it has been found that these phrase maps are not useful as it does not include the emotions of the user. A new emotional phrase map can be created such as one listed (Fig 2.1)

Fig 2.1 Emotional Phrase Map for mental health

This is a new design artifact that includes the primary and secondary emotions of the user. The last two columns have been added to create this phrase map. By using these labels it can automatically create a sub token that can be mapped to NLU/NLP system to create the emotional map. In the above diagram this phrase was uttered by a user with mental health issues and is a simulated text. This can be used in various scenarios to produce infographics for an organization, social media app etc. By deciphering the user's emotions their behavior can be analyzed and actions if needed, can be taken before it escalates. It can also be adopted to detect cyber threats targeting vulnerable individuals, mental health issues, organizational abuse, mental well being in assisted living care, domestic violence, threats etc.

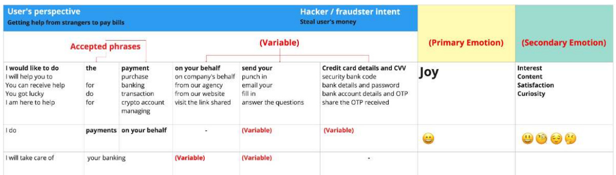

Fig 2.2 Emotional Phrase map for cyber threat towards vulnerable user

There are anomalies in emotional phrase map as it maps the user's emotions. In Fig 2.2 the user shows joy as there is a person volunteering to help with his payments. This is again a simulated text showing how a vulnerable user will react. This intent is again indeed harmful to the said user. This needs further research and classification of emotions in such situations.

The various dimensions of gathered data based on emotions can be used to provide crucial analytics to improve customer experience, emotional wellbeing, monitoring tool for assisted living etc. paving way to new products and solutions.

III. CHALLENGES

Emotional Conversational analytics: Human emotions are varied as per the situation or the personality of the individual. The sentiment analysis[13] and scoring gives a good idea about negative, positive or neutral feelings based on the keywords uttered. They are rated as per the product the users react to and utter. Deciphering primary emotions has improved tremendously with the existing NLP as it is able to classify them. But the secondary emotions need to be analyzed extensively as the primary emotions related to it might not be necessarily the same. In a conversation, a sarcasm can be deciphered differently. For example, instead of I love you the user might utter I hate you, if it is taken in a perceived sense[9] then the emotions related to it will differ.

User Intent: User intent in emotions is one of the core challenges any emotional conversational analytics platform will face. The emotions related to certain phrases might differ based on culture[11], personality of the user, native speakers, slangs[15]. This is true for any human conversation. The complexity of the conversation increases whilst spoken out of context in occasions. The result may vary depending on the audience. But deciphering such emotions based on the user intent is a challenge.

Algorithms: Neural network[14], transfer learning[12] are few of the algorithm that is suitable to decipher emotions and language processing. There is a need to increase the accuracy based on hypothetical classification of emotions. This may vary, as there is a need to set the classifications to be a common base before attempting to further sub classify secondary emotions. The local variations of user intent and emotional response needs to be evaluated in a meaningful manner if the system is to be used efficiently.

Slangs: How far can the user's private slangs hold against normal human conversations? Especially in restricted communities' humans used to substitute phrases with words that is known inside their community and is understandable. But if a system is to decipher these slangs a complex system is needed to decipher the emotional aspects of that conversation. The library then needs to add these slangs as acceptable sub phrase(s).

IV. CASE STUDY- EXAMPLE SCENARIO

For research purpose, multiple scenarios are considered to decipher user's emotions. Cyber threat, Mental wellbeing, Organizational abuse, domestic abuse was focused on this case study and corresponding emotional phrase maps were designed. The training dataset of the phrases used in this project is from Sentiment140. An online tool was built to label existing sentences on emotions.

The initial step is to label the phrases, this was done manually and in the future version once the system is trained it can be automated. The text is cleaned of emojis/slangs then phrases were labelled(Fig 4.1)

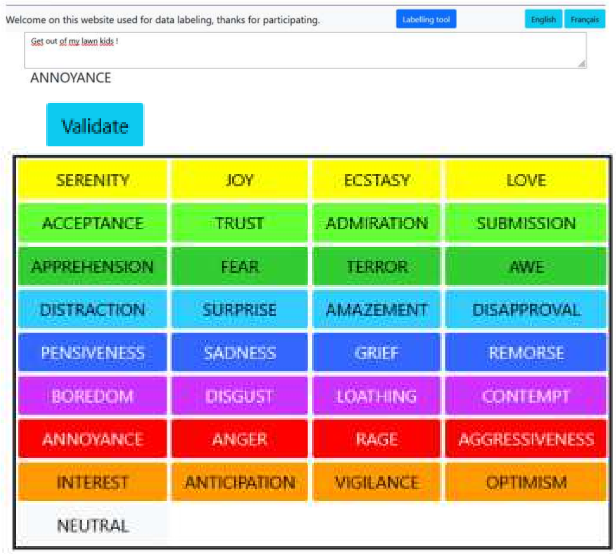

Fig 4.1 Table of primary and secondary emotions

Transfer learning algorithm was considered in this case study for layering various sentiments across the phrases. It deciphers primary and secondary emotions on the training data. Languages form word roots with structured verbs, adjectives etc. The stemming finds the word extension and removes it. Lemmatization knows the word as such and brings it back to its root via a lexicon. This reduction technique is vital whilst incorporating emotion in connection with a word sequence; certain roots and extensions carry the same emotion.

Therefore, it allows us to reduce the number of variations to its strict minimum to recognize an emotion. It also helps to reinforce the impact of a word for emotion recognition and to standardize a text in a consistent way. Example: love, loving, lover all corresponds to love. NLTK library which offers both stemming tools and lemmatization tools was chosen[3]. For the lemmatization WordNet word bag is used to assign the words to treat tags like "Adverbs" or "Pronouns" so that they receive the corresponding root. As a result, lemmatization[5] requires more preparation to be implemented than stemming which simply truncates the word and singularizes it. The 2 approaches offer quite similar results. Although lemmatization has slightly higher precision and is also increasingly preferred in NLP projects.

In this case study a generic model with 28 different labels as input was used to classify the text. This method took time to train and offered very poor results. This led to creation of several models and each one specializing on one type of emotion. Considering this, the final product is an algorithm constituting different types of layers in the model with the following architecture:

• 1st layer: “Colours” such as Neutral, Yellow, and so on (Fig 4.1)

• 2nd layer: “Sub Colours” which deals with the emotion inside of a colour for example:

If the 1st layer found a higher probability of “Yellow” then the 2nd layer will be between the 3 emotions inside such as Serenity if its probability is higher than the other 2. This layering architecture offers a possibility to tune finely each Colour’s model. In addition, Neutral colour is the most represented in our dataset and its model tended to not achieve great results. To counter this issue, few strategies were explored to improve its training quality:

1. Consider that secondary emotions are “in between” emotions, in the case 2 “Colours” has been predicted and both are neighbours (thus having a common secondary emotion), it is concluded that the secondary emotion is the right prediction. In case of 3 “Colours”, higher probability is chosen.

2. Adding secondary emotion classifier models in the 1st layer, and simply choosing the highest probability for the prediction.

The “Colours” layer contains 9 binary classifier models or 17 depending on the strategy used, which prevents emotion from overlapping. The “Sub Colours” layer contains 8 classifier models defining probabilities for 3 emotions
per model. The resultant accuracy and loss are analyzed and listed(Fig 4.2)

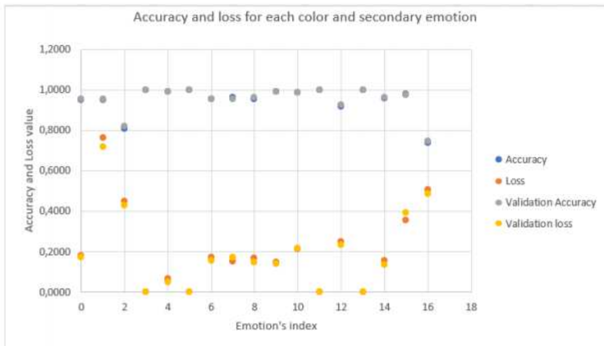

Fig 4.2 Accuracies and losses for each primary and secondary emotion(s)

IV Conclusions and Future Work

The growth of voice devices is increasing and there is a need to include human emotions into the system to address various issue to make it natural and humanlike behaviour[8]. It can also be used in various discipline to produce intelligent analytics. For example, an organizational violence happening in a region or an area of a MNC can be identified via infographics once the emotional intent of the user has been deciphered. This can be used to create smart notifications that can be used for prevention and action can be taken as per the nature of the issue.

The mental wellbeing of a patient or residents in care homes can also be determined and actions can be taken accordingly. Organizations are implementing voice bots/ robots that greet and cater to the user needs as per the user requests. The robots read facial expressions to respond in a realistic manner. As the voice over devices usage is increasing year on year, there is a need for such a system to humanize the devices.

The future of this application can be adopted to reality technologies in everyday activities and implemented to support independent living including vulnerable elderly by auto monitoring their emotions and address these needs before they escalate. Further research is needed on analysing tone to detect aggressive behaviour and other aspects of human emotions that can be incorporated into NLP and NLU. This system can be adopted to personalize voice chatbots/ virtual assistants to determine emotional needs of the user and cater accordingly.

As more and more of the users are migrating to voice over devices, it is imperative to add emotional recognition [10] into the NLP and NLU for better understanding and simulating human behaviour in systems. Especially this is prevalent with vulnerable and elderly who otherwise refrain from using technology.

These voice devices pave the way for users to access information easily without learning to use the interface that needs typing. There is a need to train the system to improve the accuracy as this research is in the early stages and requires a larger data set to determine anomalies. The accuracy rate for the case study is 24% and needs further research in providing cutting edge analytics based on the data gathered. It can also promote "humanlike" behaviour in the systems.

The wide spread of such technology can change the way/behaviour of the consumers and businesses opening to newer possibilities of deep personalization of their services and challenges to contend with.

ACKNOWLEDEMENT

I sincerely thank Prof. J.C. Nobel, Kingston University, Gaëtan Barret, Himanshu Bhasker, Shubhra Singh, Harshil Shah, Mokshi Charunika Dangampolage for their assistance.

REFERENCES

[1] Plutchik R., Kellerman H. 2013 Theories of Emotion. Volume 1 of Emotion, theory, research, and experience, Academic Press,1483270017, 9781483270012

[2] Sabharwal N., Agrawal A. 2021 Introduction to Natural Language Processing. In: Hands-on Question Answering Systems with BERT. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6664-9_1.

[3] Millstein F. 2020. Natural Language Processing With Python: Natural Language Processing Using NLTK

[4] Withanage, P., Liyanage, T., Deeyakaduwe, N., Dias, E., & Thelijjagoda, S. 2018. RoadNavigation System Using Automatic Speech Recognition (ASR) And Natural Language Processing (NLP). 2. I. (R10-HTC) (Dü.). içinde Malambe: IEEE. doi:10.1109/R10-
HTC.2018.8629859

[5] Po-Hao Chen, 2020. Essential Elements of Natural Language Processing: What the Radiologist Should Know, Academic Radiology, Volume 27, Issue 1,Pp 6-12,ISSN 1076-6332,https://doi.org/10.1016/j.acra.2019.08.010.

[6] Thejaswee M., Srilakshmi V., Anuradha K., Karuna G. 2021. Performance Analysis of Machine Learning Algorithms for Text Classification. In: Luhach A.K., Jat D.S., Bin Ghazali K.H., Gao XZ., Lingras P. (eds) Advanced Informatics for Computing Research. ICAICR 2020. Communications in Computer and Information Science, vol 1393. Springer, Singapore. https://doi.org/10.1007/978-981-16-3660-8_39

[7] Emily M. Bender, Alexander K. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics.

[8] Harari, Y. Reboot for the AI revolution. Nature 550, 324–327 (2017). https://doi.org/10.1038/550324a

[9] Xiaojuan Ma, Emily Yang, Pascale Fung. 2019. Exploring Perceived Emotional Intelligence of Personality-Driven Virtual Agents in Handling User Challenges. WWW '19: The World Wide Web Conference. https://doi.org/10.1145/3308558.3313400.

[10] Valentina Ferretti, Francesco Papaleo. 2018. Understanding others: Emotion recognition in humans and other animals, Wiley online library. https://doi.org/10.1111/gbb.12544

[11] Arvid Kappas, Nicole C. Krämer. 2011. Face-to-Face Communication over the Internet: Emotions in a Web of Culture, Language, and Technology, Cambridge University Press

[12] D. Wang and T. F. Zheng, "Transfer learning for speech and language processing," 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015, pp. 1225-1237, doi: 10.1109/APSIPA.2015.7415532

[13] R. Liu, Y. Shi, C. Ji and M. Jia, "A Survey of Sentiment Analysis Based on Transfer Learning," in IEEE Access, vol. 7, pp. 85401-85412, 2019, doi: 10.1109/ACCESS.2019.2925059.

[14] Mohamed Morchid. 2019. Neural Networks for Natural Language Processing. Artificial Intelligence [cs.AI].
Avignon Université, fftel-02398814v2f.

[15] Zhewei Sun, Richard Zemel, Yang Xu; 2021. A Computational Framework for Slang Generation. Transactions of the Association for Computational Linguistics; Vol.9 pp462–478. doi: https://doi.org/10.1162/tacl_a_00378.