**An Emotionally Aware Dialogue system with Memory**

**Vanshika Reddy**

Department of Computer Science

and Software Engineering

Rose-Hulman Institute of Technology

reddyvs@rose-hulman.edu

## **Abstract**

There has been a lot of research involving human emotions and how to interpret them. Most of this research concentrates on the identification of the response and response generation for sessions. In this paper, we propose a unique way to build a memory along with the analyzing the user’s input to generate emotionally appropriate responses. Our model contains two important and distinct features, which are generating a mental model, which acts as memory for the system and analyzing the intensity of the emotion. By remembering instances and intensities of an emotional event, out model tries to etch out an emotional profile of the user which is a key input in the response generation process. The contribution of this paper is system to better understand human emotions and provide human-like emotional assistance. (put a few data pieces here to finish the abstract)

# **I**       **Introduction**

People are complicated beings and we feel a wide range of emotions. People constantly need to express themselves and, in this era, where technology is becoming part of the mundane it is important for conversational AI to understand and respond to human emotions. Understanding human emotions can lead to increased collaboration and efficiency in HCI but for purely conversational AI, this provides the user with an assistant or a friend if one needs.

Dialogue systems such as Siri and Alexa do not do well in responding to a user’s comments that carry deep emotions. For example, when told that “my dog has just died,” the Google assistant responds with “Top ways to dispose of a dead dog's body”. When told “I had a bad day,” Siri responds with “You can always talk to me.” However, when done so, Siri breaks the flow by saying “I’m sorry, I don’t know what that means, but I can search the web for you”. Clearly there is room for improvement. We are in the process of developing a system that captures and stores emotionally charged user comments and learns from them so as to react better in the future. We note that the degree of the emotional response entirely depends on the user. We acknowledge that while one user may be in tears over the passing of their dog, another user may have a much lesser emotional response or even one of a polar opposite. Our system is designed to build a model of a user and to appropriately respond to the given user’s likely emotional response.

# **II Background**

Emotional recognition has been an ongoing area of research and has made significance progress from when it started. Classifying emotions started with humans trying to understand emotions and then building several frameworks to try and accomplish what human level of emotional understanding. Though our paper focuses more on generating a mental model of the user and identifying the intensity of an emotion, emotional analysis is the first and basic step.

A prominent field where high emotional skills are required is psychiatric counselling and there have been work done with respect to creating chatbots that can behave as counselors. One such framework proposed in 2017 [7] uses a multi-modal approach that comprises of neural nets to identify emotions based on counselling sessions, and then it generated responses based on high-level Natural language understanding using keywords and context. Parallelly, another research used social media to derive emotions from text and generate responses [3]. These responses were not specifically for healthcare but to personalize a user’s experience with the bot. A four layered Neural network was used to extract information and generate responses.

While all this work is based purely on textual analysis, another big step in understanding emotions is basing recognition on the para-lingual components of speech. Work done in this specific field also continually improves and the current model intends to use a prior work that implements a two-branch neural network structure to analyze text along with para-lingual components to detect emotions [12]. The proposed model builds onto to it by enabling the system to associate an intensity to the identified emotion.

(Mental Model introduction and more work done before)

# **III Architecture**

The basic flow of the system is as represented by figure 1. The user sends a response, which is received by the system as text and audio both. The input goes through an Emotion Recognition model that uses audio and textual cues to identify an emotion along with an intensity rating associated with the identified emotion. The input is broken down into 5 parts: Intention, emotion, context, intensity and keywords. The intention, context, keywords and the intensity are directly fed into the mental model to check for references, previous instances and update the mental model. The emotion goes through an emotion change detector before going into the mental model to extract additional information. The mental model also uses personal information to build on itself.

The last two steps are Response Generation and Response Personalization. Response Generation uses the mental model, personal information and keywords to generate a response. The generated response is similar to a generic templated response. The final step would be personalizing the response using the intention of the conversation and the keywords directly into the response before sending it back to the user.

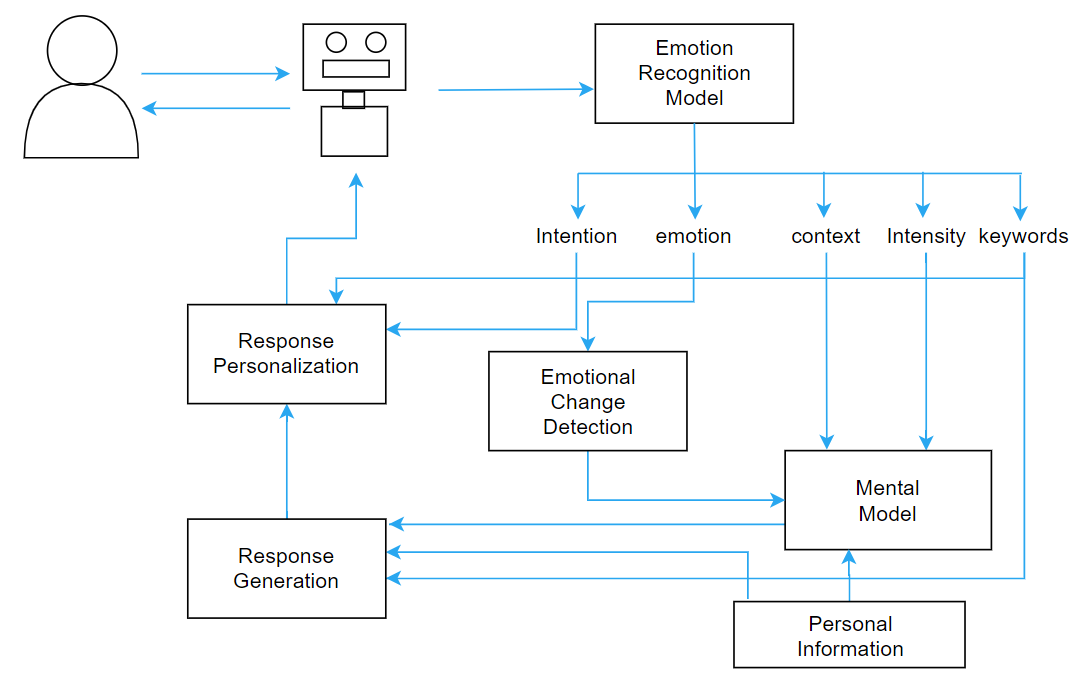


Figure 1:Proposed System Architecture and flow of system.

## **Emotion Detection**

Detecting emotions happens to be the very first step of the process and one of the most important processes of the system. It is not the core of this paper nevertheless, the architecture and system that goes behind detecting the emotions cannot be ignored.

According to prior research [9] humans feel a spectrum of emotions, but these emotions can be broadly classified into 8 emotions: ecstasy, admiration, terror, amazement, grief, loathing, rage, and vigilance as shown in figure 2.

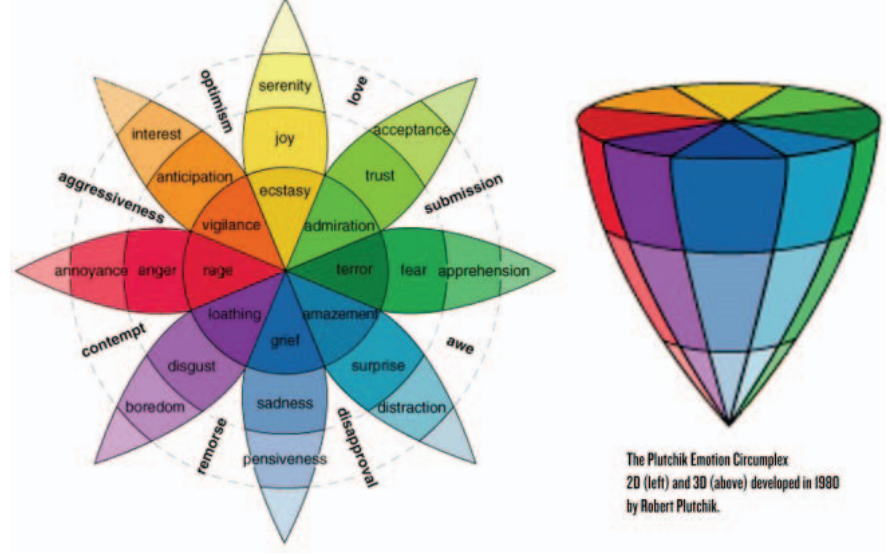


Figure 2: Plutchik Emotional Circumplex

A combination of text analysis and voice frequency is what we propose to use to detect one of the 8 emotions and the intensity of an emotion. The Intensity of an emotion spans through the scale of the single detected emotion. In this paper if the detected emotion is grief, the system would rate the intensity of the grief on a scale from 0-10, 0 being not at all feeling grief or neutral to 10 being despondent or grieving with 5 leaving the emotion of sadness.

The system architecture for the detection network is based on prior research ‘Paralinguistic Emotional Analysis with Deep Learning’ which uses a two-branch structure detecting emotions from text and voice separately. The used architecture was initially built to identify an emotion but by tweaking the system to use the frequency and other voice cues we can generate an intensity rating of the emotion based purely on audio. Once a rating is associated based on audio, any strong words detected that helped in deciding the emotion are used to finalize the intensity rating.

A list of emotions is defined with a number associated to each emotion. Every detected dialogue or instance is associated with a number that indicates what emotion it belongs to. Converting the emotions into integers makes it easier to compare emotions and understand the change in the emotions. The emotions and intensity rating would help navigate through the spectrum and understand the emotion felt by the user. As we mentioned before, the paper focuses on only one emotion so the range we cover would be 10 out of the 81 possible detectable emotional states.

The emotional state can be represented as a three-digit numerical value with the first digit indicating the emotion and the remaining two indicating the intensity of the emotion.

(expand on how the system works more after implementing Oscar’s work)

## **Emotional Change Detection**

The Emotional change detection is an additional step to help the system how the user’s emotions change with respect to events and how long an emotion persists. This step detects a change in the emotion and records the amount of time a user stays in an emotion. The change in between emotions can be gradual and shift on the scale before changing emotions completely or can be sudden depending on the person. It is important to remember that the emotional change detection only detects a change in the emotion and not the emotional state i.e. a difference between emotional state 110 and 108 would not be detected, but 110 and 208 would be detected as the emotion completely changes.

The information extracted, which includes the changed emotion, the changed emotional state and the duration of the previous emotion. The mental model records the instance with the context of why it changed. This step helps provide additional insights into the user’s mental model and helps build a more accurate one.

## **Mental Model**

The mental model is akin to a summary of the user’s mental state. The mental model tries it’s best to capture the emotional make of the user. It starts off with basic information and responses based upon a few questions that the system asks the user as a part of setup. The initial step is important to give the system an idea of how the initial responses look like and then improves on them based on the development of the mental model.

In its current state, the mental model is a series of key, value pairs with the keys being iterations of the keywords detected. A keyword is chosen if the keyword doesn’t get a close enough proximity to the existing keywords based on wordnet. For example, the vector similarity between dog and cat is above 0.8 and hence whichever keyword is seen first becomes the key and the remaining becomes a part of the value, which will be explained soon. Given a completely different word like exam that produces a lower similarity rating a new key will be created if it doesn’t fit into any existing keys.

The value associated with the keys is a linked list of instances. Each instance captures the context, exact instance, emotional state, date and time, changed emotional state, duration of emotion, a link to instance of changed emotion, and an approved response. Each context is a few main words from the instance which encompasses the instance, or all the keywords detected from the instance. When a similar key exists, the new instance is added to the existing value at the tail. An Instance reference can be under multiple keys but there is only one instance. Each individual reference has a link to the next reference pointing to a chronological sequence of instances under a key.

(include an example or diagram)

The mental model instances can be updated when an emotion change is seen or if a response becomes invalid. The mental model provides an overview of what to expect from the user’s emotional profile and helps in the response generation process by providing key pieces of information and associations. The mental model is a collection of all memories that seem important and may impact how the user responds. The simplest version only catches the immediate meaning and supports the response generation process. The mental model also behaves like a data source whose data is completely self-generated and sorted.

(Talk about more of the implementation and how the analysis on the data is done, making assumptions, drawing logic etc.)

**Response** **Generation**

(talk about this section after a little deliberation on how the data can fit in and work)

**Response** **Personalization**

(write in brevity as this diverts the focus of the paper)

# **IV Results**

(Record Calculation of Intensity accuracy and predictions. Show mental model keyword detection and association vs human rating)

# **V Conclusion and Future Work**

(Mention how the project could be expanded to other emotions similarly and how more analysis can be done on the )

# **Acknowledgements**

## **References**

1. Annabell Ho, Jeff Hancock, Adam S Miner, Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot, Journal of Communication, Volume 68, Issue 4, August 2018, Pages 712–733, <https://doi.org/10.1093/joc/jqy026>
2. Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (n.d.). Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance.
3. Dongkeon Lee, Kyo-Joong Oh and Ho-Jin Choi, "The chatbot feels you - a counseling service using emotional response generation," 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, 2017, pp. 437-440, doi: 10.1109/BIGCOMP.2017.7881752.
4. Endang Wahyu Pamungkas. (2019). Emotionally-Aware Chatbots: A Survey.  
   Ahmed Fadhil, & Gianluca Schiavo. (2019). Designing for Health Chatbots.
5. Heleen Rutjes, Martijn C. Willemsen, and Wijnand A. IJsselsteijn. 2019. Considerations on Explainable AI and Users’ Mental Models. In Where is the Human? Bridging the Gap Between AI and HCI, Workshop at CHI’19, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 5 pages. <https://doi.org/0>
6. Johnson-Laird, P. N. (2010). Mental models and human reasoning. *PNAS,* *107*, 43rd ser., 18243-18250.
7. K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375, doi: 10.1109/MDM.2017.64.
8. R. E. Banchs, "On the construction of more human-like chatbots: Affect and emotion analysis of movie dialogue data," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, 2017, pp. 1364-1367, doi: 10.1109/APSIPA.2017.8282245.
9. R. Plutchik, "Emotions and Life: Perspectives from Psychology," Biology, and Evolution, Washington, DC: American Psychological Association, 2002
10. Srivastava, T. (2018). Replicating Human Memory Structures in Neural Networks to Create Precise NLU algorithms [Web log post]. Retrieved December 10, 2020, from https://www.analyticsvidhya.com/blog/2018/04/replicating-human-memory-structures-in-neural-networks-to-create-precise-nlu-algorithms/
11. Winkler, Rainer & Söllner, Matthias: Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. 2018. - Academy of Management Annual Meeting (AOM). - Chicago, USA.
12. Youngquist, L. M. (2020). Paralinguistic Emotional Analysis with Deep Learning (Master's thesis, Rose-Hulman Institute of Technology, 2020)