Emotion Recognition in Dialogue Systems

Mehak Piplani
Disha Kedige Chandrashekarachar
Rhushabh Vaghela

Problem Statement

Conversation in its natural form is multimodal. In dialogues, we rely on others' facial expressions, vocal tonality, language, and gestures to anticipate their stance. For emotion recognition, multimodality becomes particularly important when the language is difficult to understand. In these cases, we resort to other modalities, such as prosodic and visual cues. The project aims to identify the verbal and non-verbal cues and their impact in determining the emotion of a person during a conversation.

Motivation

Our research work will be focused on studying the capability of AI to understand human emotions in real-time as inspired by the study [1]. Understanding emotions will allow the systems to adapt according to the responses and behavioral patterns. Working with artificial agents such as Cortana, Alexa, Siri, etc in our day to day life, motivated us to add the emotional cognitive ability to the responses given by artificial agents. This can be very useful for real-time personal assistants such as Siri, Google Assistant to interact more naturally and responsively while communicating with the users via both voice and text.

Relevance to course objectives

The project mainly focuses on multimodal learning of expression used as emotional receptors in generative dialogue systems. Our learning objectives will be analyzing the performance of the Emotion classes and sentiment classes by understanding how emotions are conveyed by artificial agents during a dialog. We will also analyze, recognize and predict human communicative behaviors.

Evaluation Metric

The process of emotion detection can be mapped to a multi-class classification problem. Hence, we are planning to use accuracy, F1 score, precision, and recall rates to measure the performance of our model.

Data

We are planning to utilize the dataset named Multimodal EmotionLines Dataset (MELD) [2]. It consists of audio and visual modality along with the text. It has more than 1400 dialogues and 13000 utterances from the Friends TV series. Multiple speakers participated in the dialogues. Each utterance in dialogue has been labeled by any of these seven emotions -- Anger, Disgust, Sadness, Joy, Neutral, Surprise, and Fear. MELD also has sentiment (positive, negative, and neutral) annotation for each utterance.

References

[1] Mittal, Trisha, et al. "M3er: Multiplicative multimodal emotion recognition using facial, textual, and speech cues." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 02. 2020.

[2] Poria, Soujanya, et al. "Meld: A multimodal multi-party dataset for emotion recognition in conversations." *arXiv preprint arXiv:1810.02508* (2018).