Abstract

The aim of this project is to examine emotions in three different modalities of communication and validate the presence of similar emotions in each of those modalities at any given time, and detect potential leakage of emotions. The three modalities in question include two non-verbal modalities, i.e. Facial Expressions and Voice, and one verbal modality, i.e. Speech/Text.

The extraction of emotional information from facial expressions was performed using an intelligent system, called Emotient FACET. This system makes use of Computer Vision and Machine Learning to generate emotional information based on the movement of facial muscles. Similarly, for extracting emotional information from voice, a system called OpenSMILE was used which makes use of Machine Learning and Speech Analytics to analyze speech signals. A Bag-of-Words approach with pre-categorized lexical databases has been used for analyzing the word choices of the speakers.

The dataset consists of 27 audio and videotapes of 4 politicians during public appearances such as speeches, press conferences, and interviews.

The main contribution of this project was to perform Data Fusion using the three modalities and analyzing the data to draw conclusions regarding the similarity of emotions in the three modalities. Data Fusion is the main problem and refers to the process of combining the data from the three sources that have different rates of transmission and are thereby perceived at different rates by humans.

A separate pipeline of programs was created for processing the raw output of each of the three modes of communication before fusing the processed outputs. This included carefully selecting the data, writing custom programs for processing the raw output obtained by Emotient and OpenSMILE and performing sentiment analysis on verbal data before fusing the outputs into a single file.

We can also detect potential leakage of emotions by identifying the presence, or lack thereof, of similar emotions at a given time frame from the fused data.