Report

<u>Context-aware Emotion Recognition through Posture and Speech Modulation Analysis</u>

The problem that we are going to address here includes being able to make a context-aware emotion recognition system that imposes on analyzing postures and speeches. Our project will help significantly in today's world as it will focus on Human-Computer Interactions and in understanding human emotions.

We have made our project with Python as its main language of writing and used librosa in order to get features like frequency, pitch etc. Of all the research papers that we have read about the topic we have pursued we have found that the project requires preprocessing audio files, and also extracting relevant features using Mel-Frequency Cepstral Coefficients (MFCC) technique, and training a model using a dataset of labeled audio files.
The library that we have used is the python-NLP library, SpaCy. The main reason to use it is to extract information.

The proposed method for context-aware emotion recognition through posture and speech modulation analysis involves the following steps:

Data preprocessing: The audio and visual data are preprocessed to extract relevant features. The audio features include MFCCs and prosodic features such as pitch, intensity, and duration. The visual features include body language features such as pose estimation, gesture recognition, and facial expressions.

Feature selection: The most relevant features are selected using feature selection techniques such as principal component analysis (PCA) and mutual information.

Feature fusion: The selected audio and visual features are combined using a fusion approach such as early fusion, late fusion, or hybrid fusion.

Classification: A deep learning approach such as MLP, convolutional neural network (CNN), or long short-term memory (LSTM) is used to classify emotions based on the fused features.

Evaluation: The performance of the proposed method is evaluated using metrics such as accuracy, precision, recall, and F1 score on a test dataset.

The repository contains a Jupyter notebook named "IR_NLP.ipynb" that implements a baseline emotion recognition system using speech features. The code works as follows:

Data loading: The MSP-IMPROV dataset is downloaded from a URL and extracted using the "wget" and "tar" commands.

Data preprocessing: The audio files are converted to a common format using the "sox" command. The "pyAudioAnalysis" library is used to extract features from the audio files, including pitch, intensity, and chroma.

Feature selection: The most relevant features are selected using feature selection techniques such as mutual information.

Classification: A support vector machine (SVM) classifier is trained on the selected features to classify six emotions (anger, disgust, fear, happiness, sadness, and surprise).

Evaluation: The performance of the system is evaluated using metrics such as accuracy, precision, recall, and F1 score on a test dataset.

The code in the notebook is well-commented and provides a clear explanation of each step. It also includes visualizations to help understand the data and the performance of the system.
However, it is important to note that this code only implements a baseline emotion recognition system using speech features. To improve the performance of the system, additional features such as body language features can be incorporated, and deep learning approaches can be used for classification.