Abstract

This project explores the development of a Speech Emotion Detection system, an extension of Natural Language Processing (NLP) within the broader field of Artificial Intelligence (AI). The primary objective is to create an application capable of recognizing and interpreting human emotions from speech inputs. Leveraging deep learning techniques, the system analyzes audio files to predict the emotional state conveyed in the speech.

The frontend of the application is built using React.js, providing users with an intuitive interface to upload or record audio directly. Upon submission, the audio data is sent to a Flask-based backend server. The backend handles audio processing tasks, utilizing the Librosa library to extract meaningful features such as Mel Frequency Cepstral Coefficients (MFCCs), chroma, and mel spectrogram frequencies. These features are instrumental in capturing the nuances of human speech that correlate with different emotions.

A Convolutional Neural Network (CNN) model is employed to classify the emotions present in the speech. The model is trained using the RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) dataset, which provides high-quality audio recordings of emotional speech from professional actors. This dataset ensures a diverse and representative sample of emotions, enhancing the accuracy and reliability of the emotion detection system. The model is trained on a diverse dataset encompassing various emotional expressions like happiness, sadness, anger, fear, disgust, and surprise. Data augmentation techniques, including the addition of noise and shifting, are applied to enhance the model's robustness and generalization capabilities.

The integration of the trained model into the application enables real-time emotion prediction, providing users with immediate feedback. The system not only identifies the emotional tone but also displays corresponding emojis and background effects to enhance user experience.

This project exemplifies the convergence of AI and human-computer interaction, demonstrating how machines can be taught to understand and respond to human emotions. Such a system holds significant potential for applications in fields like mental health monitoring, customer service optimization, and interactive entertainment, where emotional awareness is paramount.