Speech-Based Sentiment Analysis

Vivek Pandey(M23CSA541) and Abhisek Kumar Singh (M23CSA503)
Department of Computer Science
Indian Institute of Technology, Jodhpur
Email: m23csa541@iitj.ac.in , m23csa503@iitj.ac.in

***Abstract*—This project develops a speech-based sentiment analysis system that classifies audio signals into positive, negative, and neutral sentiments using emotion recognition. The system employs deep learning models trained on publicly available emotional speech datasets and maps detected emotions to sentiment categories. The work is organized into two phases: Phase 1 focuses on training an emotion detection model using CNN/LSTM, while Phase 2 refines sentiment mapping for improved accuracy. The analysis is conducted offline without real-time processing.**

I. Introduction

Speech sentiment analysis plays a crucial role in AI-driven applications, such as virtual assistants, mental health monitoring, and customer service. Traditional text-based sentiment analysis lacks tone and emotion context, making speech-based approaches more effective. This project focuses on developing a model that extracts emotions from speech and maps them to sentiment categories in an offline setting.

II. Proposed Methodology

The system comprises three main modules:

1) **Feature Extraction:** Converts audio signals into MFCCs and spectrograms for analysis.
2) **Emotion Recognition:** Uses CNN/LSTM models trained on labeled datasets to classify emotions.
3) **Sentiment Mapping:** Maps recognized emotions into sentiment categories (positive, negative, neutral).

The analysis is conducted offline, processing pre-recorded audio files without real-time input handling.

III. Datasets

We rely on publicly available emotional speech datasets:

- **RAVDESS:** 24 actors expressing eight emotions.
- **CREMA-D:** 91 actors expressing six emotions.
- **TESS:** Female actors speaking in seven emotional states.

IV. Evaluation Metrics

The system performance will be assessed using:

- **Accuracy:** Measures the correctness of emotion classification.
- **Confusion Matrix:** Analyzes misclassifications between emotions.
- **F1-Score:** Evaluates precision and recall of sentiment classification.

V. Project Timeline and Deliverables

A. Phase 1: Emotion Detection Model (CNN/LSTM)

- Train baseline emotion recognition model using CNN/LSTM.
- Implement feature extraction and classification of emotions.
- Conduct initial experiments with different feature extraction techniques.
- Evaluate model performance using cross-validation.
- Deliverable: Midterm report and prototype demonstration.

B. Phase 2: Sentiment Mapping

- Optimize the model for improved accuracy and robustness.
- Fine-tune the model with additional emotional datasets for better generalization.
- Implement sentiment mapping to categorize detected emotions into positive, negative, and neutral sentiment.
- Develop batch processing capabilities for analyzing multiple audio files offline.
- Conduct user testing and collect feedback for improvements.
- Deliverable: Final report with experimental results, comparative analysis, and future improvements.

VI. Conclusion

This project aims to build a robust speech sentiment analysis system for real-world applications. By integrating deep learning models and offline speech processing, it enhances AI-driven interactions in diverse fields such as healthcare, customer service, and virtual assistants.