SPEECH EMOTION RECOGNITION USING CNN

Introduction

Speech Emotion Recognition (SER) is a key area in affective computing that aims to identify and analyze emotions from spoken language. This project involves building a real-time SER system using convolutionalneural networks (CNNs) to classify emotions from audio data. The model is trained on the RAVDESS and TESS datasets and integrated into a user-friendly interface for practical applications.

Background

Emotion recognition from speech is crucial in various fields, including human-computer interaction, mental health monitoring, and customer service. Traditional methods rely on handcrafted features and classical machine learning algorithms. Recent advancements in deep learning, particularly CNNs, have significantly improved the accuracy and efficiency of SER systems. This project leverages these advancements by employinga CNN model to recognize emotions from audio recordings.

Learning Objectives

1. Understand the principles of speech emotion recognition and feature extraction.
2. Learn to preprocess and augment audio data for model training.
3. Develop skills in constructing and training CNN models for SER.
4. Evaluate the performance of the SER model using standard metrics.
5. Integrate the trained model into a real-time application with a graphical user interface (GUI).

Activities and Tasks

1. **Data Collection and Preparation**: Gather and preprocess audio data from the RAVDESS and TESSdatasets.
2. **Feature Extraction and Augmentation**: Extract Mel-Frequency Cepstral Coefficients (MFCCs) fromaudio files and augment the data to enhance model robustness.
3. **Model Development**: Construct and train a CNN model for emotion classification.
4. **Performance Evaluation**: Assess the model's performance using metrics such as accuracy and the classification report.
5. **GUI Integration**: Develop a Tkinter-based GUI for real-time emotion recognition from audio input.
6. **Model Deployment**: Save the trained model and integrate it into the GUI for practical use.

Skills and Competencies

- **Deep Learning**: Understanding of CNN architecture and its application in SER.
- **Audio Processing**: Proficiency in using librosafor feature extraction and augmentation.
- **Python Programming**: Expertise in Python, including libraries such as TensorFlow, Keras, NumPy,and Pandas.
- **GUI Development**: Experience in creating user interfaces with Tkinter.
- **Model Evaluation**: Knowledge of performance metrics such as accuracy, precision, recall, and the F1score.

Feedback and Evidence

Feedback was gathered through user testing and model performance metrics. Users found the GUI intuitive andthe system responsive. The model demonstrated high accuracy and robustness across different emotional classes. The classification report provided detailed insights into the model's performance, confirming its effectiveness.

Challenges and Solutions

1. **Data Imbalance**: Some emotions were underrepresented in the datasets.
    - **Solution**: Data augmentation techniques were used to balance the dataset and improve model generalization.
2. **Noise in Audio Data**: Background noise affected the quality of the audio samples.
    - **Solution**: Implemented noise reduction techniques and data augmentation to mitigate the impactof noise.
3. **Real-time Processing**: Ensuring the system could process and classify emotions in real-time.
    - **Solution**: Optimized the model and used efficient feature extraction methods to ensure lowlatency.

Outcomes and Impact

The project successfully developed a real-time SER system with a high degree of accuracy and responsiveness.The integration of the model into a Tkinter-based GUI made it accessible and practical for users. The system has potential applications in various domains, including mental health monitoring, customer service, and interactive voice response systems. The positive feedback from users and the detailed performance evaluation underscores the project's success.

Conclusion

This report outlines the development of a real-time SER system using CNNs with 93.4% accuracy. The project achieved its objectives, providing a powerful tool for emotion recognition from speech. The integration of themodel into a user-friendly GUI enhances its accessibility and usability. Future work could focus on expandingthe system to support additional languages and exploring its application in real-world scenarios. The project demonstrates the potential of combining deep learning models with practical interfaces to create impactful applications in affective computing