Lovish Kundu

CS767 ADVANCE MACHINE LEARNING AND NEURAL NETWORKS

Case Study on Emotion Recognition from Speech

Abstract

Problem Statement: Speech emotion recognition (SER) stands as a critical pursuit in understanding and interpreting the emotional content embedded within human speech. The project's focus was to leverage deep learning techniques for accurate and nuanced emotion recognition from audio data.

Dataset: A diverse corpus of speech datasets was amalgamated from Kaggle repositories, encompassing a rich spectrum of emotions, including joy, sadness, anger, and more. These datasets constituted the foundation for model training and evaluation.

Technology & Methodology: The project employed Convolutional Neural Networks (CNNs), specifically tailored for sequential audio data analysis. A meticulously designed CNN architecture with Conv1D layers, pooling operations, and dense layers formed the backbone for the emotion recognition model.

Uses and Benefits: The developed model aimed to decode emotional cues from speech, fostering advancements in human-computer interaction, mental health diagnostics, and personalized user experiences. The potential applications extended across domains requiring emotion-sensitive systems.

Drawbacks and Challenges: Challenges encompassed data heterogeneity, model generalization, and the nuanced nature of emotion recognition, posing hurdles in accurately capturing subtle emotional variations from speech samples.

Results: Rigorous model training and evaluation revealed promising outcomes in recognizing emotions from speech with an accuracy of 61%. However, challenges persisted in accurately discerning intricate emotional nuances, warranting further investigation and refinement.

Model Evaluation: A comprehensive analysis of the confusion matrix illuminated the model's performance across different emotion categories, providing insights into its strengths and limitations in decoding emotional cues from speech.

Conclusion: This project journeyed through data collection, preprocessing, model development, and evaluation in the pursuit of unraveling emotions from speech. While showcasing promising results, it also shed light on the complexities and challenges inherent in SER, paving the way for future advancements and refinements in emotion recognition technology.