Enhancing Speech Emotion Recognition Using Long Short-Term Memory Network

Abstract

Speech emotion recognition (SER)plays a crucial role in enhancing human-computer interaction by enabling machines to understand and respond to human emotions. In our project, we explore the limitations of the previously employed method that combines Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks for SER tasks. While this fusion approach has shown promise, it is hindered by the inherent limitation of LSTM cells, which retain the output of the CNN for a specific time instant. This characteristic is inadequate for capturing the complex temporal dependencies present in time-series data, leading to suboptimal performance in recognizing emotions. To address this drawback, we propose a novel method that leverages Three-layer stacked LSTM networks, which are designed to capture temporal features more effectively by processing the input data. This approach allows the model to utilize contextual information from time steps, enhancing its ability to discern emotional nuances in speech. We utilize the Toronto Emotional Speech Set (TESS) dataset, which provides a rich source of emotional speech data for training and evaluation. Additionally, we incorporate Mel-frequency cepstral coefficients (MFCCs) as supplementary features to enrich the representation of spoken words. MFCCs are known for their effectiveness in capturing the spectral characteristics of audio signals, which are vital for emotion recognition. Our experimental results may demonstrate that the proposed Multiple layer LSTM based method significantly outperforms the traditional CNN-LSTM fusion approach, achieving higher accuracy in classifying emotional states. This advancement not only contributes to the field of speech emotion recognition but also paves the way for more sophisticated emotion-aware systems in various applications.

Keywords: LSTM, CNN, TESS, RAVDESS, EMO-DB Speech Emotion Recognition, HCI, MFCC**.**

Batch Number: C-XVIII

Under the Guidance of:

(Signature of the Guide)

Dr K. Sravanthi, M. Tech, Ph.D.

Associate Professor

Team Members

Peddi Ajay Kumar (Y21EC138)

Yericharla Sayomi (Y21EC196)

Ramavath Mallikharjuna Naik (Y21EC146)