Abstract

Effective human-computer interaction demands efficient emotion detection from speech signals. There are many models which are proposed to detect emotion from speech signals. Suitable features, design of appropriate models and effective datasets are the main key factors which make this task successful. This thesis is motivated to design hybrid model which effectively trained to detect emotions from speech signal.

In this research, **RAVDESS** dataset is used for audio signals. The dataset contains 8 different emotions and **24** different actors, who recorder their statements in **8** different emotions. For this multiclass classification task, spectrograms of audio signals taken as input and fed to the hybrid model **CNN+LSTM+Attention.**

Current research presented a model which has four convolutional blocks. Each block contains one convolutional layer, batch normalization layer, activation layer and max pooling layer. LSTM layer is used for accumulating long term dependencies. Finally attention layer which is used to give more attention to the required emotion decisive part of speech signal to improve the performance of model. Experimental results on **RAVDESS** database shows promising results with this hybrid model. The proposed **CNN+LSTM+Attention** model is able to attain accuracy between **94% - 96.55%.**