Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name: Soma Pavan Kumar
Email: spkumar1998@gmail.com
Contribution:

1. Data Importing and Cleaning
2. Data Visualizations and getting insights
3. Getting the Statistics of Data
4. Data Modeling with Deep Learning Algorithm
    1. GRU model
    2. LSTM model

5.Conclusion

Please paste the GitHub Repo link.

GitHub Link: -
https://github.com/PavanKumar181098/Speech-Emotion-Recognition-Capstone-Project

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Write here the short summary

What is Speech Emotion Recognition?

Speech is the most natural form of communication between humans and computers. Speech is a complex signal.

it contains information regarding message speaker language and emotion. so, emotion makes speech more attractive more effective more expressive. speech emotion recognition means understanding the emotional state of a human by extracting or detecting the feature extracted by his/her voice.

There are some universal emotions including Neutral, anger, joy, sadness in which any intelligent system with limited computer resources can be trained to recognize or synthesize as needed.

In this project, we divide this project into four parts first EDA, Data Augmentation, Feature Extraction, Model.

To start with EDA first we understand our data, utilizing EDA into the bar graph, wave graph and

The Key Feature we use MFCC and Mel Spectrogram.

MFCC –

The MFCC feature extraction technique includes windowing the signal, applying the DFT, taking the log of the magnitude, and after that warping, the frequencies on a Mel scale, MFCC alone can be utilized as the feature for speech recognition. The recorded speech signals are inspected and stored utilizing Audacity.

Mel Spectrogram :

The Mel spectrogram is used to provide our models with sound information comparable to what a human would perceive. The raw sound waveforms are passed through filter banks to get the Mel spectrogram. After this process, each test includes a shape of 128 x 128, showing 128 filter banks used and 128 time steps per clip.

Now, Move towards the Second Part is Data Augmentation :

Data augmentation scheme for automatic speech recognition. that acts specifically on the spectrogram of input expressions. The augmentation policy consists of swapping blocks of frequency. channels and swapping blocks of time steps.

data augmentation scheme for automatic speech recognition. that acts specifically on the spectrogram of input expressions. The augmentation policy consists of swapping blocks of frequency. channels and swapping blocks of time steps. There are 3 fundamental ways to augment data which are time warping, frequency masking, and time masking.

Now, Feature Extraction

The reason for feature extraction is to demonstrate a speech signal by a predetermined number of components of the signal. Feature extraction is finished by changing the speech waveform to a frame of parametric representation at a generally lesser data rate for subsequent handling and examination.

Now, Model

In Model after we ran the CNN model, we get model3.h5 file and that we store it and that we used into the testing purpose for web application

Problem Statement :

Verbal Communication is effective and wanted in workplace and classroom environments alike. there's no denying the notion that Indians lack verbal communication and consequently lag behind within the workplace or classroom environments. This happens despite them having strong technical competencies. Clear and comprehensive speech is that the vital backbone of strong communication and presentation skills.

Approach:

My approach towards the Project First I understand the data, data is in a good shape or not. And my Next Aim is to convert the audio waveform into a spectrogram. Now I move towards the Augmentation part. I used Three way to Augment data Time Warping, Frequency Masking, Time Masking. The performance of the model did not improve much when I used time warping, this approach is discarded if the resource is limited. To generate syntactic data for audio, I have applied noise injection, the time has changed, pitch and speed have changed. After completing Augment part

I start LSTM part in LSTM first, the algorithm reduces the computational complexity by modifying the forgetting gate of traditional LSTM without sacrificing performance and second, in the final output of the LSTM, an attention mechanism is applied to both the time and the feature dimension to obtain the information related to the task, rather than using the output from the last iteration of the traditional algorithm.

Conclusion:

In this project, we utilize a few excellent procedures like LSTM, GRU.After utilizing all the demonstrations LSTM gave a good accuracy. After that we also use Data Augmentation could be a strategy utilized to extend the sum of information by including marginally adjusted duplicates of as of now existing information or recently made engineered information from existing information. It acts as a regularizer and makes a difference decrease overfitting when preparing a machine learning demonstration. So, this extension makes a difference to anticipate feeling utilizing discourse.