SPEECH EMOTION RECOGNITION

YAMINI S, NELOPHER NISHA M,

Dr.T.TAMILVIZHI, M.E., Ph.D.,

yaminisundar2002@gmail.com, ashraafnisha3@gmail.com

ABSTRACT

The human voice is very complex and carries multiple emotions. Emotion in speech carries insight about human actions. Through further analysis, we can better understand the motives of people.In this proposed project, we perform speech data analysis on speaker discriminated speech signals to detect the emotions of the speakers involved in the conversation.We will use Convolutional Neural Networks and LSTM to classify opposing emotions.We use statistics relating to the pitch, Mel Frequency Cepstral Coefficients (MFCCs) and Formants of speech as inputs to classification algorithms. The emotion recognition accuracy of these experiments allow us to explain which features carry the most emotional information and why. It also allows us to develop criteria to class emotions together. Using these techniques we are able to achieve high emotion recognition accuracy.

INDEX TERMS- *Mel Frequency Cepstral Coefficients, Convolutional Neural Networks, LSTM, Emotion recognition.*

RELATED WORK

EXISTING SYSTEM

The Existing system uses the supervised learning algorithm Multi-layer Perceptron which includes one or more non-linear hidden layers that leans the function

$$f(.):R^m \rightarrow R^o$$

And adopts Multi-task learning(MTL) which has a non-convex loss function where there exists more than one local minimum. Therefore different random weight initializations can lead to different validation accuracy and it requires tuning a number of hyperparameters such as the number of hidden neurons, layers, and iterations and MLP is sensitive to feature scaling.Although Multitask learning enhance the performance it fails to predict the low and extremely high scores.

PROPOSED SYSTEM

Proposed system includes the following steps

- Inputting the audio signal
- Feature extraction
- Feature enhancement
- Classifier training
- Emotion detection

The audio signal input is preprocessed before **feature extraction** to remove unwanted noise signal.The features are extracted using **MFCC** technique .Other techniques are LPC and PLP.Classification is performed which maps the features to emotion. Deep learning algorithm **CNN** is used for classifying the emotions which shows a larger learning rate and

PROPOSED ALGORITHMS

1.Convolutional Neural Networks(CNN) using Keras

2.LSTM

CNN

Convolutional Neural Network (CNN) is a type of Deep Learning architecture commonly used for image classification and recognition tasks. It consists of multiple layers, including Convolutional layers, Pooling layers, and fully connected layers.We have developed the CNN model with Keras and constructed it with 5 layers — 4 Conv1D layers followed by a Dense layer.

- Step 1: Convolution Operation
- Step 1(b): ReLU Layer
- Step 2: Pooling
- Step 3: Flattening
- Step 4: Full Connection

LSTM

1. Define Network
2. Compile Network
3. Fit Network
4. Evaluate Network
5. Make Predictions

LSTMs are a special kind of RNN — capable of learning long-term dependencies by remembering information for long periods is the default behavior.

Step 1: Decide How Much Past Data It Should Remember

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

f_t = forget gate
Decides which information to delete that is not important from previous time step

Step 2: Decide How Much This Unit Adds to the Current State

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

i_t = input gate
Determines which information to let through based on its significance in the current time step

Step 3: Decide What Part of the Current Cell State Makes It to the Output

$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

o_t = output gate
Allows the passed in information to impact the output in the current time step

INPUT

DATASET

The Ryerson Audio-Visual Database of Emotional Speech and Song (**RAVDESS**) and SAVEE dataset

LIBRARIES

- librosa==0.6.2
- matplotlib==2.2.3
- numpy==1.15.1
- pandas==0.23.4
- torchvision==0.2.1
- wave==0.0.2
- pyaudio==0.2.11

OUTPUT

Hence our project presents a new way to give the ability to machines to determine emotion with the help of the human voice. A deployment model is designed using Flask where the user can upload an audio file from their device and the system predicts the respective emotion.

OUTPUT SCREENSHOT