IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EMOTION RECOGNITION BY ANALYSING HUMAN VOICE

[1] Venkatesh P[1], [2]Mohan Kumar K, [3]Ranjith Kumar R, [4]Sankar V

[1] Assistant Professor, [2]Student, [3]Student, [4]Student

[1]Computer Science and Engineering department, [2]Computer Science and Engineering department,
[3]Computer Science and Engineering department, [4]Computer Science and Engineering department,

[1]Adthiyamaan college of engineering, Hosur, India
[2]Adthiyamaan college of engineering, Hosur, India
[3]Adthiyamaan college of engineering, Hosur, India
[4]Adthiyamaan college of engineering, Hosur, India

Abstract: Emotion recognition is that a part of speech recognition which is gaining more popularity and wish for it increases enormously. Although there are methods to know emotion using machine learning techniques, this project attempts to use deep learning and image classification method to acknowledge emotion and classify the emotion consistent with the speech signals. Our proposed model outperforms previous state-of-the-art methods in assigning data to at least one of 4 emotion categories (i.e., angry, happy, sad and neutral) when the model is applied to the RAVDESS dataset, as reflected by accuracies starting from 68.8% to 71.8%. Various datasets are investigated and explored for training emotion recognition model are explained during this paper.

1. INTRODUCTION

Today's digital era, speech signals become a mode of communication between humans and machines which is possible by various technological advancements. Speech recognition techniques with methodologies signal processing techniques made leads to Speech-to-Text (STT) technology [1] which is used mobile phones as a mode of communication. Speech Recognition is the research topic in which attempts to recognize speech. This leads to Speech Emotion Recognition (SER) growing research topic in which lots of advancements can lead to advancements in various field like automatic translation systems, machine to human interaction, used to split speech from text so on. This paper is organized as follows. Section 2 describes background information about speech recognition, emotion recognition system, applications of emotion recognition. Section 3 explains the methods of feature extraction from speech signals. Section 4 compares various speech and songs databases prepared for research. Section 5 contains various classifier algorithms for classifying speech signals. Finally, a conclusion is given in section 6.

II. BACKGROUND INFORMATION

2.1 SPEECH RECOGNITION

Speech Recognition is that the technology that deals with techniques and methodologies to acknowledge the speech from the speech signals. Various technological advancements within the field of the synthetic intelligence and signal processing techniques, recognition of emotion made simpler and possible. It is also known as "Automatic Speech Recognition". It is found that voice can be next medium for communicating with machines especially when computer-based systems. A Need for inferring emotion from spoken utterances increases exponentially. Since there is an enormous development in the field of Voice Recognition. There are many voice products has been developed like Amazon Alex, Google Home, Apple Home Pod which functions mainly on voice-based commands. It is evident that Voice will be the better medium for communicating to the machines.

2.2 EMOTION RECOGNITION

Emotion Recognition deals with the study of refining emotions, methods used for refining. Emotion can be recognized from face expressions, speech signals. Various techniques have been developed to find the emotions such as signal processing, machine learning, neural networks, and computer vision. Emotion Recognition are being studied and spreading all over the world. Emotion Recognition is gaining its popularity in research which is the key to solve many problems also makes life easier. The main need of Emotion Recognition from Speech is challenging tasks in Artificial Intelligence where speech signals is alone an input for the computer systems.

2.3 SPEECH EMOTION RECOGNITION

Speech Emotion Recognition (SER) is also used in various fields like BPO Centre and Call Centre to detect the emotion useful for identifying the happiness of the customer about the product, IVR Systems to enhance the speech interaction, to solve various language ambiguities and adaption of computer systems according to the mood and emotion of an individual. Speech Emotion Recognition is research area problem which tries to differentiate the emotion from the speech signals. Various survey state that advancement in emotion detection will make lot of systems easier and hence making a world better place to live. SER has its own application which is explained later. Emotion Recognition is the problem in ways such as emotion may differ based on the environment, culture, individual face reaction leads to ambiguous findings; speech corpus is not sufficient to accurately infer the emotion; lack of speech and song database in many languages.

2.4 APPLICATIONS OF EMOTION RECOGNITION

Emotion Recognition can be used by the psychiatrist to predict the emotion of the person by their recorded voice. In future, it can be developed as an application to predict the emotion of every humans by their recorded call voice

III. LITERATURE SURVEY

Nithya Roopa S., Prabhakaran M, Betty.P proposed the model by delivering an accuracy rate of about 35.6% is achieved from the data model for predicting the emotions.

Data Flair team proposed the model delivered an accuracy of 52.4%. By the libraries librosa, soundfile, and sklearn (among others) to build a model using an MLP Classifier. This will be able to recognize emotion from spectrum image of the sound file.

IV. PROPOSED METHODOLOGY

This section explains the proposed methods that is used and as well database used for research

4.1 EMOTION DATABASE

RAVDESS dataset, each actor has to perform 8 emotions by saying and singing two sentences and two times for each. As a result, each actor would induce 4 samples for every emotion except neutral, disgust and surprised since there's no singing data for these emotions. Each audio wave is around 4 second, the primary and last second are presumably silenced.

4.2 MFCC

Mel Frequency Cepstral Co-efficents (MFCCs) are a feature widely utilized in automatic speech and speaker recognition. Frame the signal into short frames. for every frame calculate the spectrogram estimate of the facility spectrum. Apply the Mel filter bank to the facility spectra, sum the energy in each filter. Take the logarithm of all filter bank energies. Take the DCT of the log filter bank energies. Keep DCT coefficients 2-13, discard the remainder . Loading audio data and converting it to MFCCs format is often easily done by the Python package librosa.

4.3 CNN MODEL

The convolutional neural network (CNN) is a class of deep learning neural network. CNNs represent an enormous breakthrough in image recognition. They’re most ordinarily want to analyze visual imagery and are frequently working behind the scenes in image classification. A CNN works by extracting features from images. This eliminates the necessity for manual feature extraction. The features aren't trained! They’re learned while the network trains on a group of images. This makes deep models extremely accurate for computer vision tasks. CNNs learn feature detection through tens or many hidden layers. Each layer increases the complexity of the features**.** In this CNN model with Keras are constructed with 7 layers — 6 Conv1D layers followed by a Dense layer.

V. PREPARATION OF TRAINING DATASET

All the audio clips from the RAVDESS databases are pulled out from various sessions. Using the emotion evaluation report which is given alongside the database, various wav files are labelled and categorized into seven range of emotions as mentioned earlier. Speech signals in .wav format are converted into spectrogram images within the emotion class.

VI.EXPERIEMENTAL SETUP

This section which contains explanations about experimental setup, libraries used for deep learning which helps in emotion recognition.

6.1SYSTEM SETUP

For performing the experiment I’ve used system setup consist of Core i7 6th Generation 3.7 GHz Processor, Samsung SSB of 512 GB memory space, NVIDIA GeForce GT 730 2GB GPU Card with Windows 10 installed. For deep learning I’ve used Tensor Flow 1.5 for implementing the CNN model and Tensor Board for visualizing the learning, graphs, histograms and so on.

6.2 TRAINING METHOD

All images labelled with respective emotions are prepared for training the model. The proposed CNN model was implemented using TensorFlow. The spectrogram images were generated from the RAVDESS are resized to 500 x 300. More than 400 spectrograms were generated from all the audio files in the dataset. For each emotion, Image range of about 600 for each class of emotion is collected from the corpus database. The training process was run for 700 epochs with a batch size set to 100. Initial learning rate was set to 0.01 with

a decay of 0.1 after each 10 epochs. Training data model was performed on a single NVidia GeForce GT 730 with 2 GB onboard memory. The training took around 35 minutes and the best accuracy was achieved after 100 epochs. On the training set, a loss of 0.70 was achieved, whereas 0.95 loss was recorded on the test set. An accuracy of 65.95 % was achieved per spectrogram.

VII. RESULT & ANALYSIS

An accuracy rate of about 65.6% is achieved from the data model for predicting the emotions. It is evident from the below figure that 0.96 is the highest accuracy rate achieved during validation of data.

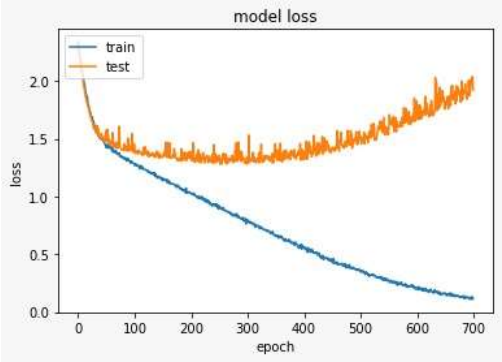

Fig1. Model loss Graph

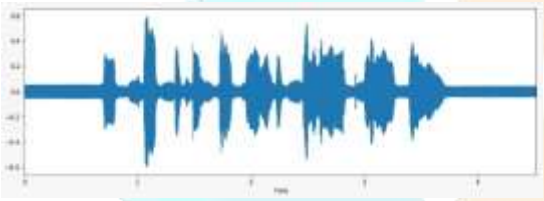

Fig 2. Audio file spectrogram image

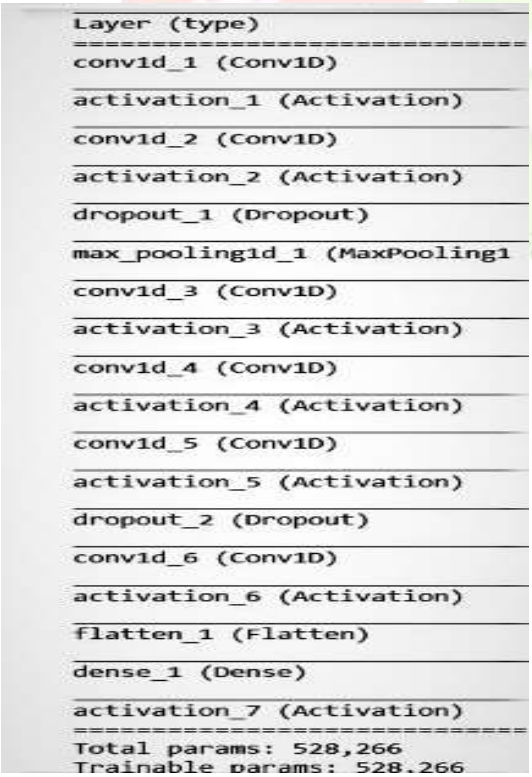

Fig 3. 7 Layers working in CNN

VII. CONCLUSION

Various investigations and surveys about Emotion Recognition, Deep learning techniques used for recognizing the emotions are performed. It is necessary in future to have a system like this with much more reliable, which has endless possibilities in all fields. This project attempted to use inception net for solving emotion recognition problem, various databases have been explored, RAVDESS database is used as dataset for carrying out my experiment. Trained my model using TensorFlow. Accuracy rate of about 66% is achieved with real time recorded audio file.

REFERENCES

[1] S. Furui, T. Kikuchi, Y. Shinnaka, and C. Hori, "Speech-to-Text and Speech-to-Speech Summarization," vol. 12, no. 4, pp. 401–408, 2004.
[2] M. El Ayadi, M. S. Kamel, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognit., vol. 44, no. 3, pp. 572–587, 2011.
[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[4] J. Schmidhuber, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015.
[5] M. T. Allen, P. A. Obrist, ''Interactions of respiratory and cardiovascular adjustments to behavioral stressors,'' Psychophysiology, vol. 23, no. 5, pp. 532–541, 1986, doi: 10.1111/j.14698986. 1986.tb00669. x.
[6] D. Carroll, J. R. Turner, and J. C. Hellawell, ''Heart rate and oxygen consumption of psychologic challenge: The effects of level of difficulty,'' Psychophysiology, vol. 23, no. 2, pp. 174–181, 1986, doi: 10.1111/j.1469-8986. 1986.tb00613. x.