International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 31st Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/83

Title **HUMANE -A VOICE BASED EMOTION PREDICTOR**

Volume 12, ISSUE 03, Pages: 577-581

Paper Authors

M. MOHANA DEEPTHI, Jaya Sudha. K, Lakshmi Jyothi. M, Bindu. M

International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

HUMANE -A Voice Based Emotion Predictor

M. MOHANA DEEPTHI[1] - Assistant Professor, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India.

Jaya Sudha. K[2], Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India. (19HP1A0573).

Lakshmi Jyothi. M[3], Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India. (20HP5A0508).

Bindu. M[4], Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India. (20HP5A0507).

Abstract

In a world like today, where days start and end with smart voice assistants, speech emotion recognition plays a major role. The development in science is incredible as it is making the human life easy and simple. Speech emotion recognition is the concept in which a machine can be able to sense the emotions of human. Instead of using hybrid classifiers, we chose only one classifier, that is, MLP (Multi-Layer Perceptron). With the help of MLP we built a machine learning model that gives above 80% accuracy. We developed an application in such a way that it takes instant inputs from end users and predicts emotion based on their voice note. Also, the sound wave of that voice note can also be seen.

Keywords— Supervised Learning, MLP-Classifier, MFCC, Model, Feature Extraction.

I. Introduction

Human can connect to someone only through emotions. But when everything in today's generation, is taking over by machines, man wants to teach machines to understand the tricky concept called "Emotions". Of course, Many scientists and researchers working rigorously on the topic of speech emotion recognition. However, making machine to learn about how humans feel is a new and at the same time, is difficult task to perform. Here we are making it simple with our application for a machine to learn and predict emotions. It is simple because we are using only one classifier to predict emotions based on the audio of various end users irrespective of their age, language, and gender with greater accuracy.

Emotion generally describes one's state of mind. Predicting emotions is difficult because people don't prefer expressing everything in the same way. The common types of emotions are happy, sad, angry, disgust, neutral and surprised. To classify these emotions from a given speech sample is the goal of this paper. We chose Multilayer Perceptron and some other methods to predict these emotions. We compared the 2 classifiers such as Support Vector Machine (SVM) and Multi-Layer Perceptron Classifier (MLP Classifier). Support Vector Machine only classifies employing a single plane and restricts the prediction. As SVM only works on one plane and thus it faces problems addressing complex statistic-based data. Also, the computational time and dependency on other algorithms are high in SVM. Whereas Multi Layer Perceptron classifier (MLP) takes less computational time and provides results with high accuracy.

II. Literature Survey

Prior research has explored Speech Emotion Recognition using various techniques and datasets, including neural networks and classifiers for emotion classification. This study presents a novel approach using MFCC to extract features from pre-processed audio files. Results show that the proposed method achieves 72% and 63% accuracy for emotion recognition using CNN and Decision Tree algorithms, respectively. The proposed methodology includes a merged deep CNN architecture that is designed in two steps, utilizing both 1D and 2D CNN architectures, which are merged to create a two-branch network. The 1D CNN is used to learn deep features from audio clips, while the 2D CNN is designed to learn deep features from log Mel Spectrograms. Transfer learning is used to speed up the training process, and hyperparameters are optimized using Bayesian optimization. In comparison, H.K. Palo used MLP networks for Emotion Recognition and extracted features such as MFCC, LPC, LPCC, and PLP from given audio signals, which were then classified using the MLP classifier.

III. Methodology

The underlying emotion in our speech is reflected in our voice through tone and pitch. During this paper we aim to classify elicit sorts of emotions like sad, happy, neutral, angry, disgust, surprised, fearful, and calm. During this paper, the emotions within the speech are predicted using neural networks. Multi-Layer Perceptron Classifier (MLP Classifier) is employed for the classification of emotions.

A. Multi-Layered Perceptron classifier and Neural Network

A Multi-Layer Perceptron (MLP) is a network which consists of three main layers such as Input Layer, Hidden Layer and Output Layer. Input Layer is used to catch the input. Output Layer is generating the response in the form of prediction and decisions from Input Layer. Hidden Layers are the layers which are present in between these input and output layers. These hidden layers vary, they often change as per the requirements.

The methodology for Multi-Layer Perceptron Network will have one input layer of (300,) and (40,80,40) hidden layers and one output layer. The input layer will take three features that are extracted from the audio file as input. The extracted features are MFCC (Mel Frequency Cepstral Coefficient), Mel, Chroma. To influence the input and to process the information the hidden layer uses activation function. This activation function uses logistic activation function. Output layer presents the output from the learned network. This layer generates the emotion as output-based computations performed by the hidden layers.

Multilayer perceptron is applied mainly for supervised learning problems. The multi-layer perceptron is used for the classification. Initially MLP is used to train the model based given Dataset. The training phase enables the MLP to seek out the correlation between the set of inputs and output. At the stage of training the MLP adjust the model parameters like weights and biases so on minimize the error.

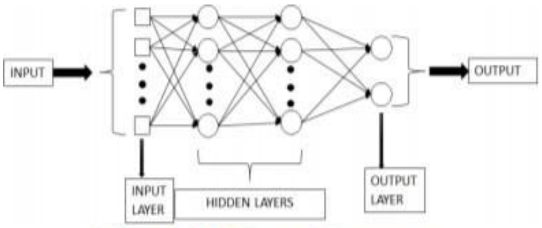

Figure.1. Multilayered Perceptron.

B. Multi-Layer Perceptron Classifier

Multi-layer Perceptron Classifier (MLP Classifier) relies on Neural Network to perform classification.

Steps involved in building of MLP classifier.

- ➢ Initialize MLP Classifier by defining and initiating the necessary parameters.
- ➢ Data is trained in Neural Network.

- To predict the output the network must be trained.
- Calculate the accuracy from predictions.

C. Features Extraction

From the input, that is based on the audio file, Feature Extraction occurs. Three main features that are extracted from the discourse signals are, MFCC, Frequency, Chroma.

MFCC

Mel Frequency Cepstral Coefficients (MFCC) is utilized to recover the sound from the given wave audio file by utilizing an HTK-styles Mel frequencies and distinct hop length.

Pitch of 1 kHz tone and 40 dB over the perceptual discernible edge is identified as 1000 mels, utilized as point of reference. The MFCC results a Discrete Cosine Change (DCT) of a real logarithm of the transient vitality shown on the Mel recurrence scale.

$$Mel(f) = 2595 \text{ x } \log 10\ (1+f/700)$$

Mel

The Mel scale relates evident repeat or pitch of a tone to its real assessed recurrence. Individuals are incredibly improved at perceiving little changes such as low frequency than they are at high frequencies pitch. This scale makes our features arrange even more eagerly so that an individual can listen.

$$M(f)=1125Ln(1+f/700)$$

Chroma

Chroma is one of the feature extractions used in this model. Chroma plays a vital role for high-level semantic analysis. In high-level tasks, to get better results we use extracted chroma features.

$$Cf(b) = \sum Z-|z=0|X1f\ (b + z\beta)\ ||$$

IV. Implementation

We developed our application with Jupiter, an open-source IDE that allows for scientific programming in Python across various operating systems. Jupiter provides a robust platform for developers to build and test applications with a range of useful features. It is user-friendly and includes editing, interactive testing, debugging, and introspection features. Using MLP Classifier we were able to build a machine learning model that gives us an accuracy of 84.5%. Just by using one classifier we made it possible to gain more than 80% accuracy. We developed an application in a way that it takes instant inputs from users and predicts emotion based on the voice note. The sound wave of the voice note is also generated.

The input will be taken through microphone, and the output will be displayed on the screen.

Figure 2. A simple Micro Phone

V. Results and Analysis

We used MLP as a Supervised Machine Learning model algorithm. Also, we mainly focused on the computing time and accuracy. Here, Machine learning model will extract the features that are present in given Dataset. Based on the extracted features, the trained model will classify emotions and predict the output. The model will again be tested with the test sample. We prefer users to give their instant input i.e., voice note through microphone to generate an emotion. The waveform for the voice note is additionally generated.

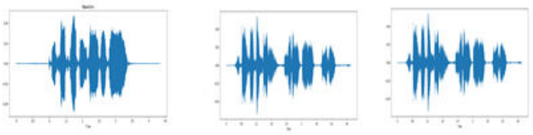

Figure3, Figure4. Sound wave-plots of test sample

In general, to describe the performance of a model, we use validation methodologies such as accuracy report, **Confusion Matrix** and accuracy score. Based on the data set a **Figure5** confusion matrix used to represent the true values in a tabular form to classify the performance. In our application, we are using this confusion matrix to check the accuracy of the classified emotions.

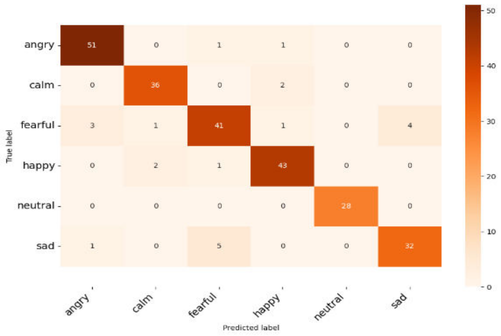

Figure5. Confusion Matrix

VI. Conclusion

Emotion plays a huge role in daily interpersonal human interactions. It helps us to match and understand the emotions of others by conveying our feelings and giving feedback to others. Emotional displays convey considerable information about the mental state of a person. Speech Emotion Recognition (SER) can recognize the emotional aspects of speech no matter what the content is. While humans can efficiently perform this task naturally, making the system to do this is ongoing subject of research. Signal processing unit is the important issue in speech emotion recognition systems because feature extraction depends on it. MLP classifier is used to predict emotions from the provided input. After extracting effective features of the given input, the classified emotion will be displayed as an output. Feature selection techniques indicate the elimination of more information in the execution, which is good in Machine Learning Application.

VII. References

[1]. S. Lalitha, A. Madhavan, B. Bhushan, and S. Saketh," *Speech Emotion Recognition*",2014 International Conference on Advances in Electronics Computers and Communication, 2014, pp.1-4, Doi:10.1109/ICAECC.2014.7002390

[2]. In their 2015 article "Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000 to 2011," Anagnostopoulos, Iliou, and Giannoukos explore various features and classifiers used for speech emotion recognition during the 11-year period. The article was published in the Artificial Intelligence Review journal and covers pages 155-177 in volume 43, issue 2.

[3]. S. Lalitha, Shikha Tripathi, "Emotion detection using perceptual based speech features," *India Conference (INDICON) 2016 IEEE Annual, pp. 1-5, 2016.*

[4]. Mingke Xu, Fan Zhang, Samee U. Khan, "Improve Accuracy of Speech Emotion Recognition with Attention Head Fusion", *Computing and Communication Workshop and Conference (CCWC) 2020 10th Annual, pp. 1058-1064, 2020.*

[5]. Dellaert, F., Polzin, T. and Waibel, "*A Recognizing emotion in speech*". In the Proceedings of ICSLP 3, (Philadelphia, PA, 1996).

[6]. The Artificial Intelligence Review journal published an article in 2015 by Anagnostopoulos, Iliou, and Giannoukos titled "*Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000*

to 2011." The article delves into various features and classifiers that were utilized for speech emotion recognition during the period of 2000 to 2011, and it spans across pages 155-177 in volume 43, issue 2.

[7]. K.S. Rao, T.P. Kumar, K. Anusha, B. Leela, I. Bhavana, Gowtham S.V.S.K. "*Emotion Recognition from Speech*" (IJCSIT), 2012.