Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 6, June 2019, pg.161 – 164

EMOTION DETECTION USING MACHINE LEARNING

Animesh Nema[1]; Nishant Ragate[2]; Avaneesh Pratap Singh[3]; Asst. Prof. Shashikala H K[4]

[1]Department of CSE, School of Engineering and Technology-Jain University, Bangalore, India
[2]Department of CSE, School of Engineering and Technology-Jain University, Bangalore, India
[3]Department of CSE, School of Engineering and Technology-Jain University, Bangalore, India
[4]Department of CSE, School of Engineering and Technology-Jain University, Bangalore, India
[1] animesh.nema23@gmail.com; [2] nragate1@gmail.com; [3] avaneeshsingh242@gmail.com;
[4] shashi.hk85@gmail.com

Abstract— Programmed Speech feeling acknowledgment has been a consuming issue since a decade ago, analysts have been endeavouring to build up a framework progressively like human, for feeling acknowledgment. Discourse has numerous parameters which have extraordinary weightage in perceiving feeling to be specific prosodic and spectral highlights, out of prosodic highlights to be specific pitch, intensity and energy are famously utilized and out of spectral highlights formant Mel Frequency Cepstral Coefficients are normally utilized by scientists around the world. Further the classifiers are prepared by utilizing these highlights for ordering feelings precisely, this venture is an endeavour to improve the existing innovation to achieve higher exactness and more extensive scope of feeling acknowledgment from discourse utilizing idea of Mel-recurrence Cepstrum Coefficients (MFCC) in the Python (Jupyter Notebook) An effective feeling acknowledgment framework can be valuable in the field of restorative science, mechanical autonomy building, call focus application and so forth.

Keywords— Machine learning, Continuous neural network, Mel Frequency Cepstrum Coefficient, Dataset, Mood detection

I. INTRODUCTION

Human machine collaboration is generally utilized these days in numerous applications. One of the vehicles of connection is speech. The fundamental difficulties in human machine communication is recognition of feeling from speech. At the point when two people interface to one another they can without much of a stretch perceive the basic feeling in the speech expressed by the other individual. The target of feeling acknowledgment framework is to emulate the human recognition instruments. There are a few applications in speech feeling acknowledgment.

Feeling can assume a significant job in basic leadership. Feeling can be identified from various physiological sign additionally If feeling can be perceived appropriately from speech then a framework can act as needs be. A proficient feeling acknowledgment framework can be helpful in the field of medicinal science, mechanical autonomy designing, call focus application and so on.

II. LITERATURE SURVEY

The rapid development of technologies has lead in modelling of various system to perceive human emotion. There are many research happened on emotion using voice of a person.

- Title of paper: Emotion Recognition On Speech Signals Using Machine Learning.

Authors: Mohan Ghai, Shamit Lal, Shivam Dugga l and Shrey Manik

Advantages: Three classification algorithms, namely Random Decision Forest, SVM and Gradient Boosting classified an audio signal into one of the 7 classes. Out of the three, Random Decision Forest achieved the highest accuracy of 81.05%.

Disadvantages: The paper presents only the analyses of seven human emotions using speech signals. It can be expanded to predict more human emotions

- Title of paper: Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation

Authors: Thurid Vogt, Elisabeth Andr´e, and Johannes Wagne

Advantages: Speech emotion recognition system consists of three principal parts : signal processing (digitalization, filtering, segmenting the input signal into meaningful units), feature calculation and classification. Audio segmentation can be performed by voice activity detection which is a fast segmentation method not requiring high-level linguistic knowledge.

Disadvantages: The speech database used to train the classifier should be adjusted to the particular application as much as possible

III. METHODOLOGY

This project aims to build a system through which we can get the speech as input from the user and apply suitable machine learning algorithms to recognize the mood. A mic will be used to record the speech and to detect different emotion inputs.

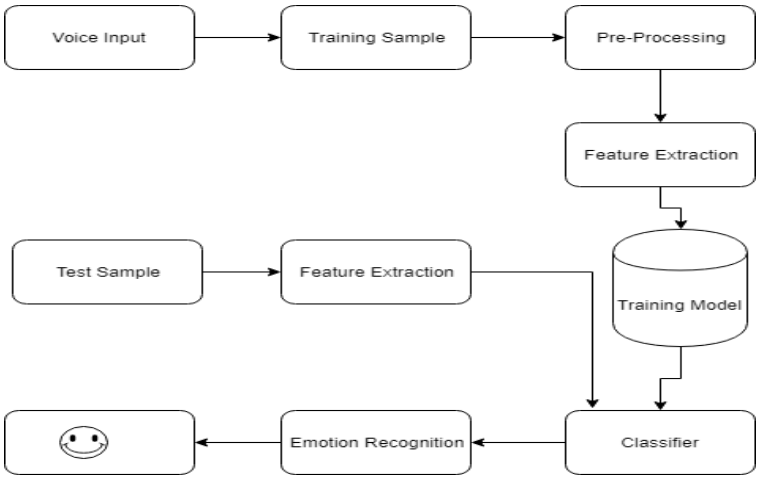

Fig. 2 Example of an unacceptable low-resolution image

A. Voice dataset:

We have used different dataset to train the machine so it can predict the most likely output. Dataset that are used are:-

- RAVDESS. This dataset incorporates around 1500 sound record contribution from 24 distinct on-screen characters. 12 male and 12 female where these entertainers record short sounds in 8 unique feelings neutral, calm, happy, sad, angry, fearful, disgust, surprised.

- SAVEE. This dataset contains around 500 sound documents recorded by 4 diverse male entertainers. The initial two characters of the record name compare to the various feelings that they portray.

B. Preprocessing

Voice data are then filtered and unwanted noise are being removed so that we can get the best model without any errors.

A. Feature extraction

Initially dataset will be in the ".wav" type files first we need to get the signal of voice so that we can get some value and distinguish uniquely all the voice samples. Each, ".wav" file will be having different wavelength, frequency, amplitude and other attributes too. This step is the important process of emotion detection since this feature will be added to the train the model.

B. Training Model

We used CNN model to solve this problem, MLP and LSTM are not that suitable for the emotion detection. CNN model was the best for our classification problem. After training numerous models we got the best validation accuracy of 60% with 18 layers, softmax activation function, rmsprop activation function, batch size of 32 and 1000 epochs.

C. Emotion Classification

Based on the trained model we will take an input from a user, it can be from mic or recorded data. Then we will find the values that are being used in the model. Predicting emotions on the test data, generating graphs for input, comparing Mel Frequency Cepstrum Coefficient of input waveform with the model. Based on the value we get model will decide and gives the output as emotion. This is the final step in emotion detection.

IV. RESULT

Predicting emotions on the test data, generating graph for input, comparing Mel Frequency Cepstrum Coefficient of input waveform with the model

```
Out[62]: array([[0.09288605, 0.09451584, 0.11181492, 0.07753757, 0.12014433,
                 0.09385058, 0.15594684, 0.14601302, 0.0571673 , 0.05012359]],
               dtype=float32)

In [63]: livepreds1=livepreds.argmax(axis=1)
         livepreds1

Out[63]: array([6])

In [64]: liveabc = livepreds1.astype(int).flatten()

In [65]: # livepredictions = (lb.inverse_transform((liveabc)))
         livepredictions = liveabc
         numpy.set_printoptions(precision=3)
         print (livepredictions)

         [6]

In [ ]:

In [ ]:
```


In this case label 6 is returned which indicates that the voice is of a male and the mood detected is calm. Hence, using this project we can successfully categorize five moods namely

1. Fearful
2. Happy
3. Calm
4. Angry
5. Sad

These moods can be detected for both male and female, and as the training set grows the accuracy of detection increases.

V. CONCLUSION

To conclude, our implementation of Mood Detection Using Speech Patterns can be used as an added service to other software systems. Building the model was a demanding task as it included part of trail and mistake strategies, tuning and so forth. The model is very much made to recognize male and female voices and it recognizes with 100% precision. The model was tuned to identify feelings with over 70% precision. Precision can be expanded by including more sample data for preparing.

We compare the accuracy of emotion by trying to implement the model using MLP,LSTM and CNN. We built a Multi Perceptron model, LSTM model and CNN models. The MLP and LSTM were not suitable as it gave us low accuracy. As our project is a classification problem where were categorize the different emotions, CNN worked best for us.

CNN model was the best for our classification problem. After training numerous models we got the best validation accuracy of 60% with 18 layers, softmax activation function, rmsprop activation function, batch size of 32 and 1000 epochs.

ACKNOWLEDGEMENT

This author would like to express their gratitude towards Department of Computer Science and Engineering for their sincere help, support and encouragement.

REFERENCES

[1] Thurid Vogt, Elisabeth André, and Johannes Wagne: Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation, ISBN:978-3-540-85098-40.
[2] Pierre-Yves Oudeyer: The production and recognition of emotions in speech:features and algorithms, Sony CSL Paris, 6, rue Amyot, 75005 Paris, France Received 28 September 2001; received in revised form 26 November 2002; accepted 30 November 2002.
[3] Assel Davletcharovaa, Sherin Sugathanb, Bibia Abrahamc, Alex Pappachen Jamesa: Detection and Analysis of Emotion From Speech Signals, 1877-0509 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.
[4] Sherif Yacoub, Steve Simske, Xiaofan Lin, John Burns: Recognition of Emotions in Interactive Voice Response Systems, To be presented at Eurospeech 2003, 8th European Conference on Speech Communication and Technology, 1-4 September 2003, Geneva, Switzerland.

