International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Emotion Recognition in voice using Deep Neural Network

CH SIVASANKAR [1], K SREE LATHA [2], G VIGNESH [3], C JASWANTH REDDY [4],
B GOPI VIVEK [5]

[1]Assitant professor, [2,3,4,5] Students, Dept of CSIT

[1,2,3,4,5] Siddharth Institute of Engineering & Technology, Puttur-517583

Abstract - In mortal machine interface operation, emotion recognition from the speech signal has been exploration content since numerous times. feelings play an extremely important part in mortal internal life. It's a medium of expression of one's perspective or bone 's internal state to others. Speech Emotion Recognition (SER) can be defined as birth of the emotional state of the speaker from his or her speech signal. There are many universal feelings including Neutral, wrathfulness, Happiness, Sadness etc., in which any intelligent system with finite computational coffers can be trained to identify or synthesize as needed. In this work, we're rooting Mel- frequency cepstral portions(MFCC), Chromogram, Mel gauged spectrogram in confluence with Spectral discrepancy and Tonal Centroid features. Deep Neural Network is used to categorize the emotion in this works.

Keywords: Speech Processing, Emotion Recognition, Deep Neural Network, SER (Speech Emotion Recognition).

Introduction

There are numerous ways of communication but the speech signal is one of the fastest and most natural styles of dispatches between humans. thus, the speech can be the fast and effective system of commerce between mortal and the machine as well. Humans have the natural capability to use all their available senses for maximum mindfulness of the entered communication. Through all the available senses people actually smell the emotional state of their communication mate. The emotional discovery is natural for humans but it's veritably delicate task for machine. thus, the purpose of emotion recognition system is to use emotion affiliated knowledge in such a way that mortal machine communication will be bettered. In this system, the quality of point birth directly affected the delicacy of speech emotion recognition. In the process of point birth, it generally took the whole emotion judgment as units for point rooting, and birth contents were four aspects of emotion speech, which were several aural characteristics of time construction, breadth construction, abecedarian frequency construction, and formant construction. also, discrepancy emotion speech with no emotion judgment from these four aspects, acquiring the law of emotional signal

distribution, also classify emotion speech according to the law.

Still, so far, no exploration on deep neural network has been applied to speech emotion processing. We set up that the DNN in speech emotion processing has a huge advantage. thus, this paper proposed a system to realize the emotional features automatically uprooted from the audio using the librosa package in python.

We used DNN to train a 5- subcaste-deep network to prize speech emotion features. It incorporates the speech emotion features of further successive frames, to make a high latitude characteristic, and uses softmax classifier subcaste to classify the emotional speech. The speech emotion recognition test delicacy reached73.38 which is a high value compared to the other models of this size. Traditional machine literacy styles are k- nearest neighbors (KNN), Hidden Markov Model (HMM) and Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian fusions Model (GMM) etc., to classify the feelings. The important issues in speech emotion recognition system are the signal processing unit in which applicable features are uprooted from available speech signal and another is a classifier which recognizes feelings from the speech signal.

The average delicacy of the utmost of the classifiers for speaker independent system is lower than that for the speaker dependent. Automatic emotion recognitions from the mortal speech are adding now a day because it results in the better relations between mortal-machine.

SCOPE:

• Automatic emotion recognition using speech can help associations to understand their guests more when in a call.

• Call center can make separate strategies on dealing with people with different people.

• For E-Learning, seminaries can cover the feelings of their scholars to more prepare their education system for the betterment of the scholars.

• Robotics has wide use of Emotion discovery as a robot designed to interact with human should understand the human's emotion.

• Emotion discovery is the key to Human Computer Interaction (HCI).

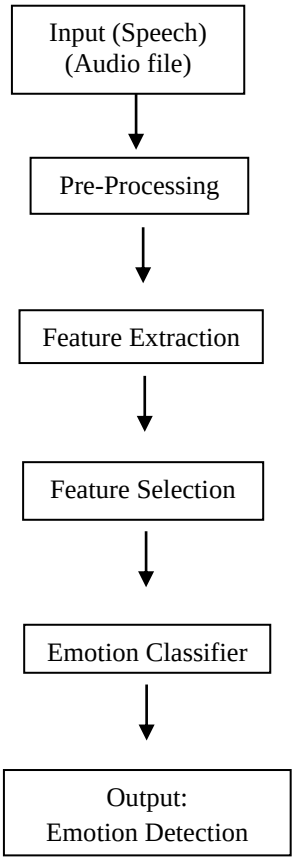
Related Work

Emotion Recognition from Speech Signal.

Emotion recognition is a fleetly growing exploration sphere in recent times. Unlike humans, machines warrant the capacities to perceive and show heartstrings. But mortal- computer commerce can be bettered by automated heartstrings recognition, thereby reducing the need of mortal intervention. In this paper, four introductory heartstrings (outrage, Happy, Fear and Neutral) are analysed from emotional speech signals. Signal recovering styles are used for carrying the product features from these signals.

2. Model Architecture

flow for the project is given below:



2. Algorithm

Convolutional Neural Network

A convolutional neural network (CNN) uses a variation of the multilayer perceptron. A CNN contains one or further than one convolutional layer. These layers can either be fully connected or pooled. Before passing the result to the coming subcaste, the convolutional subcaste uses a convolutional operation on the input.

Due to this convolutional operation, the network can be important deeper but with important smaller parameters. Due to this capability, convolutional neural networks show veritably effective results in image and videotape recognition, natural language processing, and recommender systems. Convolutional neural networks also show great results in semantic parsing and translation discovery.

Procedure:

• Upload the dataset of audio (. wav lines) to be read using librosa library.

• Data Preprocessing is a fashion that's used to convert the raw data into a clean data set.

• Drawing the data refers to removing the null values, filling the null values with meaningful value, removing indistinguishable values, removing outliers, removing unwanted attributes.

• If dataset contains any categorical records means convert those categorical variables to numerical values. The uprooted features are Mel- frequency cepstral portions (MFCC), Chromogram, Mel gauged spectrogram in confluence with Spectral discrepancy and Tonal Centroid features.

• We resolve our dataset of 1440 audio lines in 2 corridor, training data with 1008 audio lines and testing data with 432 audio lines. Then 70 of the data is taken for the training dataset.

• Deep literacy can give increased delicacy and drop in computational power.

• Deep Neural Network (DNN) is extensively used in deep literacy to train models for tasks which traditional machine learning algorithms cannot do or is hard to do.

• An audio is uploaded by the stoner (which includes speech of a person), and the model is used to prognosticate the emotion of the speaker in the audio.

• A Beaker armature grounded web operation is developed to use the model. It has 2 corridors, the system and the stoner.

• There is a stoner enrollment and login operation system in the UI.

• A new stoner first needs to register their details which includes name, dispatch and the word. This stoner information is stored in MySQL database.

• A formerly registered stoner, whose data is stored in the system's MySQL database can login to the web app using their valid credentials.

• Once they successfully log in, only also they're handed access to the operation to prognosticate the feelings.

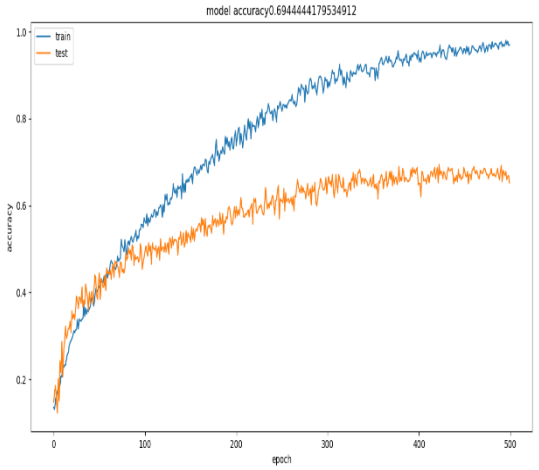
• A successional () model with 5 layers is created.

• Train dataset is used for training the dataset using 700 ages.

• The stylish one with respect to test delicacy is the model which we will use for vaticination.

• In the affair subcaste, we use softmax for bracket of the feelings.

• Softmax takes in a vector of figures and converts them to chances which are also used for image generating results.



• Softmax converts logits into chances by taking the expounders from every affair and also homogenize each of these figures by the sum of similar expounders, similar that the entire affair vector adds up to one.

4. Result:

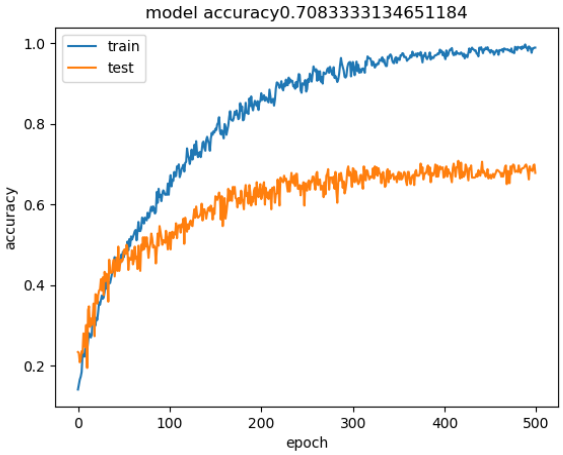
• The proposed scheme presented an approach to fete the emotion from the mortal speech.

• This approach has been enforced by the using the neural networks.

• We've successfully developed a deep literacy model using the deep neural network armature to prognosticate the feelings of the speaker in an audio.

• We've famed our design in a web grounded operation using the Flask armature. The UI also includes stoner enrollment system.

• We were suitable to get a test delicacy of 73.4 %

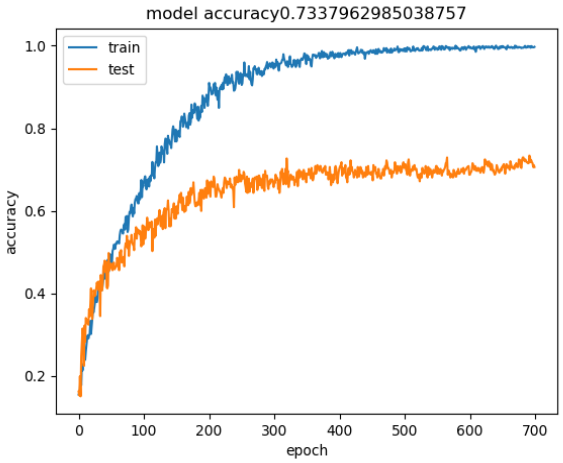


using the trained model.

• The train vs test rigor of our model over 700 ages are shown below.

• The model gets the stylish test delicacy of 73.4%

• We preliminarily tried other models too, with different rigor.



• And this is another model which we developed using DNN.

5. Conclusion:

The proposed scheme presented an approach to fete the emotion from the mortal speech. This approach has been enforced by the using the neural networks.

We've successfully developed a deep literacy model using the deep neural network armature to prognosticate the feelings of the speaker in an audio. We've famed our design in a web- grounded operation using the Flask armature.

The UI also includes stoner enrollment system. We were suitable to get a test delicacy of 73.4 % using the trained model. Deep literacy systems are suitable to induce good results. We've successfully developed a deep literacy model with 73.4% test delicacy in emotion recognition.

Please note that emotion vaticination is private and the feelings rated by a person for the same audio can differ from person to person. This is also the reason why the algorithm which is trained on mortal rated feelings can induce erratic results occasionally.

The model was trained of RAVDESS dataset, so the accentuation of the speaker can also lead to erratic results as the model is trained on North American accentuation database.

Please note that emotion vaticination is private and the feelings rated by a person for the same audio can differ from person to person. This is also the reason why the algorithm which is trained on mortal rated feelings can induce erratic results occasionally.

The model was trained of RAVDESS dataset, so the accentuation of the speaker can also lead to erratic

results as the model is only trained on North American accentuation database.

6. References:

[1] Szegedy, Christian & Toshev, Alexander & Erhan, Dumitru. (2013). Deep Neural Networks for Object Discovery. 1- 9.

[2] AshishB. Ingale &D.S. Chaudhari (2012). Speech Emotion Recognition. International Journal of Soft Computing and Engineering (IJSCE) ISSN 2231-2307, Volume- 2 Issue- 1, March 2012

[3] M.E. Ayadi,M.S. Kamel,F. Karray, " check on Speech Emotion Recognition Features, Bracket Schemes, and Databases ", Pattern Recognition 44,PP.572- 587, 2011.

[4] T. Vogt,E. Andre andJ. Wagner, "Automatic Recognition of feelings from Speech A review of the literature and recommendations for practical consummation", LNCS 4868, PP.75- 91, 2008.

[5] S. Emerich,E. Lupu,A. Apatean, " feelings Recognitions by Speech and Facial Expressions Analysis ", 17th European Signal Processing Conference, 2009.

[6] Esther Ramdinmawii, Abhijit Mohanta and Vinay Kumar Mittal, Emotion Recognition from Speech Signal. Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[7] Ruhul Amin Khalil, Edward Jones, Mohammad Inayatullah Babar, Tariquillqh Jan, Mohammad Haseeb Zafar, and Thamer Alhussain, Speech Emotion Recognition using Deep learning Techniques, 2019

[8] Pavol Harár, Radim Burget and Malay Kishore Dutta, Speech Emotion Recognition with Deep Learning, 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN).

[9] Taiba Majid Wani, Teddy Surya Gunawan, (Senior Member, IEEE), Syed Asif Ahmad Qadri, Mira Kartiwi, (Member, IEEE), And Eliathamby Ambikairajah, (Senior Member, IEEE), A Comprehensive review of speech emotion recognition system, Date of publication March 22, 2021, date of current version April 1, 2021.

[10] Jiaxin Ye, Xincheng Wen, Yujie Wei, Yong Xu, Kunhong Liu, Hongming Shan, Temporal Modelling Matters: A Novel Temporal Emotional Modelling Approach for Speech Emotion Recognition, Xiamen University, Nov 2022