



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 8, August 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Speech Emotion Recognition in Python Using Deep Learning

KM Pratistha¹, Atulya Kumari², Prof Prakash Devale³

Dept. of Information Technology, Bharati Vidyapeeth College of Engineering, Pune, Maharashtra, India^{1,2,3}

ABSTRACT: At present, speech emotion recognition was an emerging and crossing field of artificial intelligence and artificial psychology. Besides this it was a hot research topic of signal processing and pattern recognition. The research was widely applied in many fields like human-computer interaction, interactive teaching, entertainment, security fields, and so on. The basic functionality to handle all the audio file consist of human voice and generate speech, is extreme features which make humans different from other living beings. The human speech can be differentiated by some properties. In this paper we detect the basic emotion of human with the help of Stochastic gradient descent algorithm, ANN model and neural network i.e. MLP classifiers. As we know that today 95% communication is based on vocals which shows different characteristics of human and emotion is one of them which shows like fear, anxiety, happiness, sadness and angry etc. Hence, voice and speech analysis causes emotion inside human and very beneficial in different areas of communication like when a victim has been interviewed, he start stammering if he tells lie and in this case we detect victim's lie with the help of emotionanalysis.

KEYWORDS: Emotion, rate, stochastic gradient descent, neural network (MLP classifiers) and ANN model.

I. INTRODUCTION

In today's digital world, speech signals become the fastest and growing communication mode between humans and machines which is possible by various advancements of new technology. Speech recognition techniques with method of searching signal processing techniques made leads to various technology which is used by mobile phones as a greatest mode of communication. It is the proliferate research topic which set out to recognize speech signals. This leads to Speech Emotion Recognition (SER) expanding research topic in which lots of development can lead to betterment in various field like automatic translation systems, machine to human interaction and also used to integrate speech from text. In contrast the paper focus to survey and review various speech extraction features, emotional speech databases, classifier algorithms and so on. Problems present in various topics were inscribe. Speech Recognition is the automation that deals with techniques and program to recognize the speech from the speech signals. It is also known as Automatic Speech Recognition". This is investigating that voice can be next medium for communicating with machines especially for computer-based systems. A Need for inferring emotion from spoken statement increases exponentially. Since there is an extensive development in the field of Voice Recognition. Amazon Alexa, Google Home, Apple Home Pod are various voice product which are developed and functions mainly on voice-based commands. Primarily, Emotion Recognition share out with the study of inferring emotions. Emotion can be acknowledged from facial declaration and speech signals. Various techniques have been enlarged to find the emotions such as signal processing, machine learning, neural networks, computer sight. Emotion analysis, Emotion Recognition are being considered and developed all over the world. Emotion Recognition is procuring its popularity in research which is the clue to answer many issues and also makes our life trouble-free. The main require of Emotion Recognition from Speech is demanding tasks in Artificial Intelligence where speech signals is solitary an input for the computer systems. Speech Emotion Recognition (SER) is also used in varied department like BPO Centre and Call Centre to detect the emotion useful for pinpoint the happiness of the customer about the product, to solve various language vagueness and modification of computer systems according to the mood and emotion of an individual. Speech Emotion Recognition is research area problem which tries to deduce the emotion from the speech signals. Various survey state that development in emotion observation will make lot of systems easier and hence making a world superior place to live. Emotion Recognition is the challenging problem in process such as emotion may differ root on the environment, culture and separate face reaction leads to ambiguous findings, deficiency of speech database in many accents. Emotion Recognition set out as the performance variable for idiomatic analysis thus identifying the unsatisfied customer, customer satisfaction and so on. SER is used in-car board system based on the guidance of the ore state of the driver that can be provided to the system to initiate welfare and put a stop to any accident.

II. LITERATURE SURVEY

In past years, there are various type of strategy that is used for emotion intuition through speech. The elite task of emotion recognition with the help of speech is emotion modelling. The variety of classifier established such as Gaussian Matrix Model (GMM), hidden Markov Model (HMM), support Vector Machine (SVM) and many more. [5] On 5th may 2010 B. Yanguand M. Luggar proposed some class of tasks where the emotion is detected by the euphony property and that was based on music hypothesis. they can evaluate those features in more pragmatic aspect. [6] Yash pal singhchavan, M.L.dhore and Palavi Yeshaware, they proposed the speech features such as Mel Frequency Cepstrum coefficient (MFCC). speech of people comment was the input here. Support vector machine (SVM) is the classifier used for the categorizing of emotion. [7] Mina hamidi and Muharram Mansoorizade describe some attempt from sustained Persian speech with regards to the automatic recognition of emotional situation. On any account of so much shortage of significant database in that language for emotional recognition in the first place they build a database of emotional speech in Persian. The database hold 2400 wave clips modulated with anger, sad, happy and fear class of emotions. After that they withdraw prosodic features acting as pitch, intensity etc and after all this finally they appeal neural network. [9] Alessandro delfino offers the emotional state of a person which was able to determine by the system. Then it started the enrollment of audio signals. The system comprises of two functional structure gender recognition (GR) and emotion recognition (ER). For the recognition of gender pitch frequency estimation is used whereas for detection of emotion sum is used. Here, they use two SVM's particularly one for male emotion recognition and further particularly for female emotion recognition. [10] Sirisha Devi, Y. Srinivas and Shiv Prasad Nandyala established and enhance text dependent speaker recognition to expose the emotion of the speaker using hybrid FFBN and GMM methods.

III. PROPOSED SYSTEM

We will use MFCC, Chroma and Mel Frequency Cepstrum as speech features rather than raw waveform which may contain unnecessary information that doesn't help on the classification. Feature extraction is used to withdraw the feature from the speech signal. Mel Frequency Cepstral Coefficient is the most prime and expanding method for the feature extraction Mel Frequency Cepstral Coefficient (MFCC) are the features that used mostly in studies about SER. MFCC has a simple calculation and good ability of distinction and anti-noise. In MLP some neuron uses a non-linear activation function that was developed to the model. Here is the formula $Y(v_i) = \tanh(v_i)$ and $y(v_i) = (1 + e^{-v_i})$ Linear Prediction Coefficients (LPC) are also extracted from speech, serving as an alternative of short-term spectral features for comparison. In LPC gives the description about the quality of a distinct channel of any alone person and this channel characteristic will get a transformation by the various emotions, so using these features can extract the emotions in speech. First, we need to install some dependencies using pip like librosa, numpy, sound file, scikit-learn, pediocin in this diagram, we mentioned that firstly the speech is divided into various types of segments which works simultaneously. After dividing in various segments these are extracted with the MFCC features. Different features are done by different extraction process in the segments which are classified from the predefined database present over there. And after all the cycle process we get to know the emotion of the speech. We allow only these emotions like angry, sad, neutral and happy.

Database creation: -

Preparing the Dataset: Here, we download and convert our own speech, the dataset to be suited for extraction. Loading the Dataset: This process is about loading the dataset in Python which involves extracting audio features, such as obtaining different features such as power, pitch and vocal tract configuration from the speech signal, we will use librosa library to do that. Let's define a dictionary to hold numbers and emotions available in the dataset, and a list to hold those we want- calm, happy, sad, fearful etc. We'll use the glob () function from the glob module to get all path names for the sound files in our dataset. Time to split the dataset into training and testing sets. For this we use load data function for this. Testing the Model: Measuring how good our model is doing by giving some audio and see what they show as an output. Now, let's initialize an MLP classifier. This is a multilayer perceptron classifier; it optimizes the log loss function using stochastic gradient descent algorithm. MLP classifiers has an internal neural network for the purpose of classification. This is a feed forward ANN model.

Applications:

SER has entered so many areas these years, including:

- **Customer service:** In call center conversation may be used to analyze behavioral study of call attendants with

the customers which helps to improve the quality of service.

- **Recommender systems:** Can be useful to recommend products to customers based on their emotion towards that product.
- **Medical field:** in the world of telemedicine where patients are evaluated over mobile platforms, the ability for a medical professional to discern what the patient is feeling can be useful in the healing process.

IV. CONCLUSION AND FUTURE ENHANCEMENT

The proposed scheme presented an approach to recognize the emotion from the human speech. This approach has been applied by the neural network. This argumentation directs attention on the feature extraction method that is applied in the emotion recognition through speech. For the motive of feature extraction **Stochastic gradient descent algorithm is used**. The neural network is used for the training to get the better performance. The earned output represents single emotion from the given speech samples can be truly identified by the proposed system. The performance is highly hanging on the emotional speech specimen. So, it is needful to take genuine and accurate speech samples. One more planned enhancement is that positive and negative can be applied for the bigger set of emotions and can be implemented by other classified algorithms too. Firstly, the limit number of public available databases leads researchers to use a significant small number of databases to evaluate their views. The trained data and speech acquired from libraries are too small and so the method must be more vigorous in order to get a better result. Although it gives quite good result on small database. Every classifier has positive as well as negative views against the others. From the review of previous research done, it was proven that the recognition rate depends only on the features, data and classification method. Still There are some more hybrid features that were not studied yet.

REFERENCES

- [1] Ayadi, Kamel, M. S. and Karray, Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 2011
- [2] Sehepoch and Wongthanavas, Speech Emotion Recognition Using Support Vector Machines In Knowledge and Smart Technology (KST), 5th International Conference IEEE, 2013
- [3] P Harar, R Burget, MK Dutta, "Speech emotion recognition with deep learning" 4th International conference, 2017
- [4] RA Khalil, E Jones, MI Babar, T Jan, MHZafar "speech emotion techniques using deep learning" IEEE 2019 - ieeexplore.ieee.org
- [5] B. Yang, M. lugar "Emotion Recognition from speech signals using new technologies" ISSN vol 90, 5 may 2010.
- [6] Yash pal singhchavan, M.L.Dhore, pallaviyehsaware "Speech emotion using support vector machine" international journal of computer applications (0975-8887) vol- 1.
- [7] Mina hamidi and Muharram mansoorizade "emotion recognition from Persian speech with neural network" international journal of artificial intelligence and application (IJAI), vol 3, Sept. 2012.
- [8] Alam, M. J., Attabi, Y., Dumouchel, P., Kenny, P., & O. Shaughnessy, D. D. (2013). Amplitude modulation feature for emotion recognition from speech. In INTER SPEECH (pp. 2420–2424).
- [9] Alesandro del fino, Mario marchese and Andera scirone "Gender driven speech recognition through speech signals for intelligence applications" IEEE volumeno.2, Dec 2013.
- [10] J. Sirisha devi, Y. srinivas and shiv Prasad nandyalla "Automatic speech emotion & speaker recognition based on hybrid Gaussian Mixture Model" International journal on computational science and application (IJCSA) vol.4, no. 1, February 2014.
- [11] Athabi and Dumouchel "Emotion recognition from speech: WOC-NN and class-interaction" 11th international conference on information science, signal processing and their applications (ISSPA) (pp. 126–131). IEEE 2012.
- [12] E. Ayadi, M. Kamal and K array, F. (2011). Survey on speech emotion recognition i.e. Features, classification and databases. *Pattern Recognition*, 573–587.
- [13] Ankur Sapra, Nikhil Panwar, Sohan Panwar, "Emotion recognition form speech ", *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 2, Feb. 2013, ISSN 2250-2459.
- [14] D. Morrison, L. Da Silva, "Ensemble method for spoken emotion recognition on call-centres", *Speech Communication*, vol. 49, no. 2, pp. 92-112, 2007.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details