International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume28 Issue 06 – JUNE 2021.

EMOTION DETECTION USING VOICE

Mr. Kotte Sandeep[#1], Pasula Jahnavi[*2], Putchakayala Lineesha[*3], Palasa Maheswari[*4]

 # Assistant Professor, Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada, Andhra Pradesh, India.

** Final year B. Tech Undergraduate Student, Department of CSE, Dhanekula Institute of Engineering and Technology*

** Final year B. Tech Undergraduate Student, Department of CSE, Dhanekula Institute of Engineering and Technology*

** Final year B. Tech Undergraduate Student, Department of CSE, Dhanekula Institute of Engineering and Technology*

***Abstract*— Emotion detection is a new research era in health informatics and forensic technology. Voice based emotion recognition is getting popular, as the situation where the facial image is not available, the voice is the only way to detect the emotional or psychiatric condition of a person. Here, our problem is to identify how the person emotion is detected using the voice. In our project, Multi-Layer Perceptron (MLP) will be used as voice feature and a fixed valued k-means clustered method will be used for feature classification. In our project emotional conditions like happy, angry and sad etc.. Emotions will be detected from female and male voices.**

***Index Terms*— Emotion Analysis, Emotion Detection, Speech Processing.**

I. INTRODUCTION

Emotions are psychological states that result from neurophysiological changes, and have various implications for thoughts, feelings, and behavioral responses and, to some extent, pleasure or displeasure. There is currently no scientific consensus on the definition. There are many things that are intertwined with emotions, moods, temperament, personality, temperament, creativity, motivation. Emotions are complex. There are various theories as to whether emotions change our behavior. On the other hand, the physiology of emotions is closely related to the arousal of the nervous system.[1] Emotions are also related to behavioral tendencies. Extroverts tend to be sociable and expressive, while introverts are socially reclusive loners and tend to hide their emotions. Emotions are often the driving force of stimulation. Emotions, on the other hand, are not causal, but simply a syndrome of components that can include motivation, feelings, behaviors and physiological changes, but none of these components are emotional. Emotions have no substance that causes these components.[2]

Research on emotions has increased over the past two decades, with contributions from several disciplines including psychology, medicine, history, sociology of emotions, and computer science. Enthusiastic research on this subject is being driven by numerous theories that attempt to explain the origins, functions, and other aspects of emotions. Current research areas in the concept of emotions include the development of materials that stimulate and elicit emotions. Also, PET scans and fMRI scans help to study the process of emotional painting in the brain.[3]

There are different ways to recognize and detect the emotion of a person, Facial Emotion Recognition which uses a camera and tries understand the face features while an emotion is projected on the face.[4] Emotions can also be detected by calculating the Neural activity of a person inside the brain, for each and every action done there will be neural activity going inside the brain and can be recorded using Electroencephalogram also called as EEG and by calculating the activity we can recognize the emotions. Emotions can also be detected by analyzing the voice patterns spoken by the person.[5]

Humans can assess the emotion of another person by observing facial, voice but a machine cannot understand or recognize the emotion of a person. By using Natural Language Processing, we can calculate the negativity and positivity of the statement made by the person but it cannot recognize the emotion. We need an efficient and reliable way to recognize an emotion.[6]

Hardware Specification:

- Processor - Intel Corei3
- Processor Speed - 3.20 GHz
- RAM - 4 GB
- Hard Disk - 10 GB
- Monitor – SVGA
- Keyboard - Standard keyboard

Software Specification:

- Operating System - Windows
- Coding Language - Python Programming
- IDE - Jupyter Notebook
- Technology - Machine Learning

II. EXISTING SYSTEM

Among various physiological signal acquisition methods for the study of human brain, EEG (Electroencephalography) is more effective. EEG provides a convenient, non-intrusive and accurate way of capturing brain signals in multiple channels at fine temporal resolution. We propose an ensemble learning algorithm for automatically computing the most discriminative subset of EEG channels for internal emotion recognition.[7] Our method describes an EEG channel using kernel-based representations computed from the training EEG recordings. For ensemble learning, we formulate a graph embedding linear discriminant objective function using the kernel representations. The objective function is efficiently solved via sparse non-negative Principal Component Analysis (PCA)[8] and the final classifier is learned using the sparse projection coefficients. Our algorithm is useful in reducing the amount of data while improving computational efficiency and classification accuracy at the same time. Experiments on publicly available EEG dataset demonstrate the superiority of the proposed algorithm over the compared methods.

Disadvantages:

- Emotion is classified using the EEG device, which is not available to all personal and is not cost effective.
- In existing system, the system can will classify based on the neural activity in the brain, sensors will give inappropriate data some times.

III. PROPOSED SYSTEM

Because using this technology involves deep studying and analysis, it is extremely difficult for predicting the emotion. Our system helps understand the emotions and doesn’t need any facial image. Multi-layer perceptron is used as voice features and a fixed k-means clustered method will be used for feature classification. We will predict the basic emotions like Happy, Sad, Angry etc.,[9]

Advantages:

- Prediction is done on the voice from the user.
- It doesn’t need any facial images for prediction of emotion.
- Cepstral coefficient is used as a voice feature for efficacy in prediction.

IV. ALGORITHM

MULTI-LAYER PERCEPTRON:

A multilayer perceptron (MLP) is a class of feed forward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to any feed forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation); see § Terminology. Multilayer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

Activation function:-If a multilayer perceptron has a linear activation function in all neurons, that is, a linear function that maps the weighted inputs to the output of each neuron, then linear algebra shows that any number of layers can be reduced to a two-layer input-output model. In MLPs some neurons use a nonlinear activation function that was developed to model the frequency of action potentials, or firing, of biological neurons.

The two historically common activation functions are both sigmoid, and are described by

$$y(v_i) = \tanh(v_i) \text{ and } y(v_i) = (1 + e^{-v_i})^{-1}$$

In recent developments of deep learning the rectifier linear unit (ReLU) is more frequently used as one of the possible ways to overcome the numerical problems related to the sigmoids.

The first is a hyperbolic tangent that ranges from -1 to 1, while the other is the logistic function, which is similar in shape but ranges from 0 to 1. Here yi is the output of the ith node (neuron) vi is the weighted sum of the input connections. Alternative activation functions have been proposed, including the rectifier and soft plus functions. More specialized activation functions include radial basis functions (used in radial basis networks, another class of supervised neural network models).

Layers:-The MLP consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes. Since MLPs are fully connected, each node in one layer connects with a certain weight wij to every node in the following layer.

Learning:-Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result. This is an example of supervised learning, and is carried out through back propagation, a generalization of the least mean squares algorithm in the linear perceptron.

We can represent the degree of error in an output node j in the nth data point (training example) by ej(n) = dj(n)-yj(n) where d is the target value and yis the value produced by the perceptron. The node weights can then be adjusted based on corrections that minimize the error in the entire output, given by

$$\mathcal{E}(n) = \frac{1}{2}\sum_{j} e_j^2(n)$$

Using gradient descent, the change in each weight is

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n)$$

where yi is the output of the previous neuron and n (eta) is the learning rate, which is selected to ensure that the weights quickly converge to a response, without oscillations. The derivative to be calculated depends on the induced local field vj, which itself varies. It is easy to prove that for an output node this derivative can be simplified to

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n)\phi'(v_j(n))$$

where phi is the derivative of the activation function described above, which itself does not vary. The analysis is more difficult for the change in weights to a hidden node, but it can be shown that the relevant derivative is

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = \phi'(v_j(n)) \sum_k -\frac{\partial \mathcal{E}(n)}{\partial v_k(n)} w_{kj}(n)$$

This depends on the change in weights of the kth nodes, which represent the output layer. So, to change the hidden layer weights, the output layer weights change according to the derivative of the activation function, and so this algorithm represents a back propagation of the activation function.[10]

The term "multilayer perceptron" does not refer to a single perceptron that has multiple layers. Rather, it contains many perceptrons that are organized into layers. An alternative is "multilayer perceptron network". Moreover, MLP "perceptrons" are not perceptrons in the strictest possible sense. True perceptrons are formally a special case of artificial neurons that use a threshold activation function such as the Heaviside step function. MLP perceptrons can employ arbitrary activation functions. A true perceptron performs binary classification, an MLP neuron is free to either perform classification or regression, depending upon its activation function.

MLPs are useful in research for their ability to solve problems stochastically, which often allows approximate solutions for extremely complex problems like fitness approximation. MLPs were a popular machine learning solution in the 1980s, finding applications in diverse fields such as speech recognition, image recognition, and machine translation software, but thereafter faced strong competition from much simpler (and related) support vector machines. Interest in back propagation networks returned due to the successes of deep learning.[11]

V. Results

Fit the model with the training data which will the train the initialized multilayer perceptron with the training data

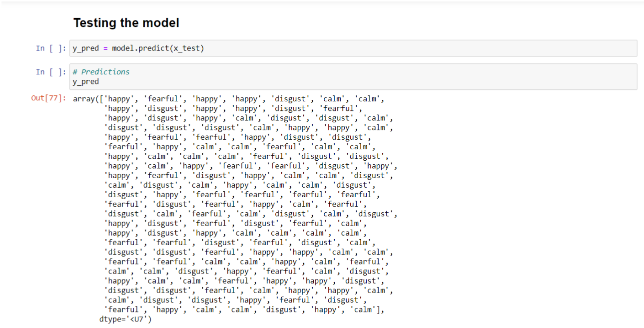

Fig. 1: Testing the Model

Test model after fitting the model with the training data, predict the test data this will return the predicted classes for the test data

Fig. 2: Final Accuracy

After testing the model with the test data, get the metrics from the multilayer perceptron and print the Accuracy metric.

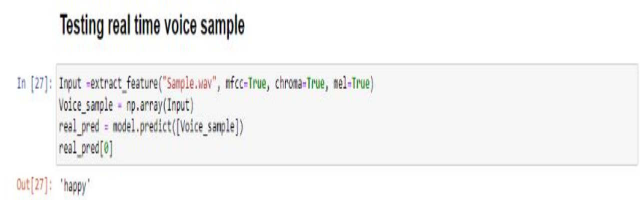

Fig. 3: Testing real time voice Sample

VI. Conclusion

The main objective of the project is to detect the emotion of the person for the psychiatric analysis and for understanding of another person and we achieved the prediction of emotion with 79.22% accuracy and we have used multi-layer perceptron as voice feature and derived each emotion with the help of the dataset and we achieved predicting emotions like happy, calm, disgust, fearful with the best accuracy. We can predict emotion by using a recorded sample of a person speaking. Our system can help people to detect and study the emotions and behavior relationships and psychiatric evaluation can also be done.

Acknowledgment

We would like to take this opportunity to express our deep gratitude to the members who assisted us directly and indirectly for the completion of this project work.

We would like to express our immeasurable sense of gratitude to Mr. KOTTE SANDEEP, Assistant Professor the project guide for his esteemed guidance and support provided to us at every step of this project work.

We would like to express our appreciation to Mr. KOTTE SANDEEP, Assistant Professor and Project Coordinator for allowing us to make this project a successful one. We would also like to earnestly thank the Project Review Committee members.

We wholeheartedly express our gratitude to Dr. SURESH SUNDARADASU, Professor & Head of the Department of Computer Science & Engineering, and to Dr. RAVI KADIYALA, Principal for the help and valuable suggestions to accomplish this project work.

We also extend our thanks to all the faculty members of the Department of Computer Science & Engineering and the Library of Dhanekula Institute of Engineering and Technology for imparting knowledge to us throughout this project.

We would like to extend our warm appreciation to all our friends for sharing their knowledge and valuable contributions to this project.

Finally, we express our deep sense of gratitude to our parents for their continuous support throughout our academic career and their encouragement in completing this project successfully.

REFERENCES

[1] Habib Ullah, Muhammad Uzair, Arif Mahmood, Mohib Ullah4, Sultan DaudKhan, FaouziAlaya Cheikh "Internal Emotion Classification Using EEG Signal with Sparse DiscriminativeEnsemble" published inIEEAccessin 2019.

[2] Seunghyun Yoon, Seokhyun Byun, and Kyomin Jung "Multimodal Speech Emotion Recognition Using Audio And Text" published in arxiv on October10th, 2018.

[3] DzmitryBahdanau, Kyunghyun Cho, and YoshuaBengio,"Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473,2014.

[4] AharonSatt, Shai Rozenberg, and Ron Hoory, "Efficient emotion recognition from speech using deep learning on spectrograms," Proc. Interspeech 2017, pp. 1089–1093, 2017.

[5] AsselDavletcharova, SherinSugathan, Bibia Abraham, Alex Pappachen James "Detection and Analysis of Emotion from Speech Signals", 2015.

[6] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 4, pp. 570-578, July 1993.

[7] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34-39, Jan. 1959.

[8] C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, and M. Miller, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767-782, May 2001.

[9] AsselDavletcharova, SherinSugathan, Bibia Abraham, Alex Pappachen James "Detection and Analysis of Emotion from Speech Signals", 2015.

[10] C.-H. Chen, W.-P. Lee, J.-Y. Huang, "Tracking and recognizing emotions in short text messages from online chatting services," Information Processing & Management, vol. 54, no. 6, pp. 1325–1344, 2018.

Mr. Kotte Sandeep: Assistant Professor, Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology

Pasula Jahnavi: Final year B. Tech Undergraduate Student, Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology

Putchakayala Lineesha: Final year B. Tech Undergraduate Student, Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology

Palasa Maheswari: Final year B. Tech Undergraduate Student, Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology