International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Speech Emotion Recognition Using Neural Network

Ritu Chaudhary, Shreya Saraswat, Siddhant Chaturvedi, Prof.Prajakta Naregalkar

(Department of Eletronics Engineering) *Bharati Vidyapeeth College of Engineering, Pune*

ABSTRACT – There are plenty of methods for emotion recognition such as facial expression recognition, gesture recognition, speech recognition. Moreover, models have been developed with the fusion of two or more of these techniques.

In this paper, we have introduced a model for emotion recognition using speech. Speech emotion recognition is considered to be of real importance in the study of psychology. SER tells about the emotion of a person using speech/voice. According to science the voice of a person can tell a lot about the underlying emotions. This paper proposes a model based on a convolution neural network (CNN) which focuses on providing effective interaction between humans and computers. This model allows a computer to understand the emotions of a human with the help of his audio sample. In the proposed model, at first, we used a random forest classifier, and after that CNN. As per the results obtained, CNN provides an accuracy of 95%.

Key words: Speech Emotion Recognition (SER), Convolution Neural Network (CNN), Random Forest (RF).

1.INTRODUCTION

Emotion classification is defined as one of the most challenging and difficult tasks in a speech signal processing domain. The problem of speaker or speech recognition can be made easier by recognizing emotion using speech. The sound signal is one of the main sources of communication and by processing it we can recognize the speaker, speech, or even emotion. The maximum simple precept defining emotion reputation indicates with studying the acoustic distinction that happens while uttering the identical factor beneath extraordinary. In addition to the features corresponding to the speaker and/or the speech, for representing the emotional state of the speaker, sound signals consist of some features. The trouble of the emotion category for human speech is addressed through the paper. The project is geared towards exploring dependencies the character of utterance has with the human emotional state. The nervous system and heart rate are directly influenced by emotions. So the emotional repute of someone also can be decided with the aid of using measuring the coronary heart rate of someone. The speech indicators also are a consultant of the coronary heart price of the speaker because the coronary heart price additionally impacts the speech. The researches show that if there may be a negative stimulus that reasons negative emotion the coronary heart charge decelerates extra actively than while there are good The numerous broadly used packages utilized in speech sign processing consists of non-touch clinical diagnosis, faraway remark of patients, human-computer interaction (HCI), It can also be analyzed to detect heart rate, which will naturally enable remote diagnosis of a patient by only using audio It is also used in medical applications like the diagnosis of the nasopharynx and vocal tract for any abnormalities. In recent years, the researcher's efforts in Human-computer interaction are the emphasis on how to make computers more sensible which can understand human emotions. The primary communication medium for interaction is speech whereas emotion in speech is useful as it does not alter the linguistic content of speech but changes its decision making, perception, cognition, creativity, attention, reasoning, and memory are straightly affected by the emotions in When we are concentrating on the emotions in speech one needs to make the following analysis: -

(i) The emotional feeling towards the hearer of the speaker. (ii) The emotional feeling towards the message of the speaker. (iii) Speakers' emotional character. The design of more natural human-machine interfaces is complementary to the proper detection of emotion from speech. To consummate for task ending jobs, the speaker's emotion can be oppressed by the system's dialog administrator to render more simple responses. Speech processing can

be divided into speaker identification, speech recognition, and emotion detection. Emotion detection is the part of speech processing in which information about the emotional content of the speaker can be retrieved. The emotional analysis will be helpful for getting the psychological condition of the person who is speaking. Knowing the same, one can communicate more effectively than that of, without knowing the same. It can also lead to effective communication between man and machine if the machine is capable of understanding the emotions of the person in the interaction. Speech recognition can find the exact words spoken by the person. It needs a huge vocabulary which will help for getting the exact words being spoken.

2.SYSTEM OVERVIEW

The implemented system consists of an Esp 8266 which is used to provide Wi-Fi connectivity to the proposed system. LEDs of different colors are connected to the Esp which thereby tells about the extracted emotion from the applied audio. These LEDs are connected with the help of connecting wires. Let us now discuss every component used in the model in detail.

Esp 8266: The ESP8266 is a cheap and user-friendly device that helps in providing internet connectivity to a project. The module can paintings each as a station (can hook up with Wi-Fi) and as an Access point (can create hotspot), consequently, it may effortlessly fetch statistics and add it to the net making the Internet of Things as clean as possible. Another interesting characteristic of this module is that it may be programmed the usage of the Arduino IDE which makes it plenty greater person-friendly. However this version of the module has only 2 GPIO pins (you can hack it to use up to 4) so we have to use it along with another microcontroller like Arduino, else we can look onto the more standalone ESP-12 It also can fetch facts from the net the use of API's subsequently your challenge may want to get right of entry to any facts this is to be had at the net, as a result making it smarter.

Features

- Low cost, compact and powerful Wi-Fi Module
- Power Supply: +3.3V only
- Current Consumption: 100mA
- I/O Voltage: 3.6V (max)
- I/O source current: 12mA (max)
- Built-in low energy 32-bit MCU @ 80MHz
- 512kB Flash Memory
- Can be used as Station or Access Point or both combined
- Supports Deep sleep (<10uA)
- Supports serial communication hence compatible with many development platforms like Arduino
- Can be programmed using Arduino IDE or AT-commands or Lua Script

Applications

- IoT Projects
- Access Point Portals
- Wireless Data logging
- Smart Home Automation
- Learn the basics of networking
- Portable Electronics
- Smart bulbs and Sockets

LED: LED is a Light-emitting diode. An electric cutting-edge passes through a microchip, which illuminates the tiny mild reasserts we name LEDs and the end result is seen mild. To save you overall performance issues, the heat LEDs produce is absorbed right into a heat sink. LED lights products produce light about 90% extra effectively than incandescent mild bulbs.

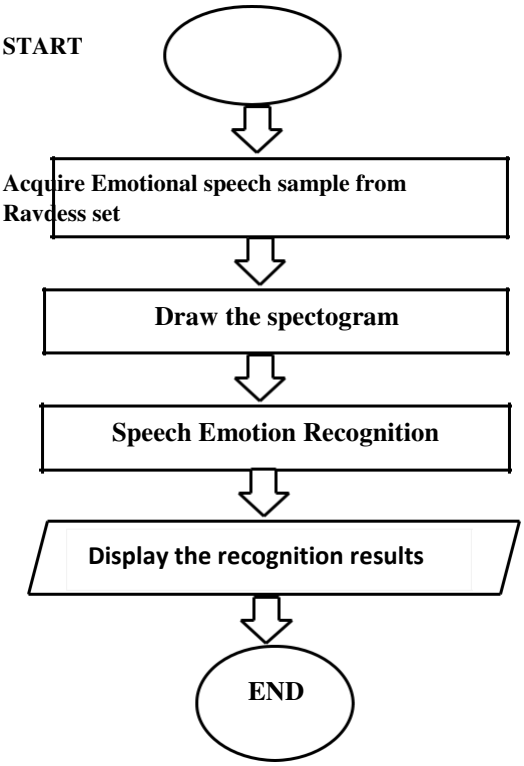
Colour code :-

1) Neutral – yellow
2) Calm- blue
3) Happy- 2 blue
4) Sad- yellow red
5) Angry- 2 red
6) Fearful- green yellow
7) Disgust- green
8) Surprised- 2 yellow

Connecting wires: Connecting wires offer a medium to an electrical modern in order for current to travel from one factor of a circuit to another. In the case of computers, wires are embedded into circuit boards to keep pulses of electricity. Most wires in computer systems and digital additives are made from copper or aluminum due to the fact copper is reasonably-priced and electrically conductive. In a primary circuit, the cord comes from one terminal of a strength source, then connects to a transfer that determines whether or not the circuit is open or closed. The related wires of a tool are used to attract energy and strength and carry out detailed tasks. Before the current can journey through the wire, the circuit needs to be closed; in different words, there can't be any breaks withinside the path. Electricity can not effortlessly journey through the air, and if it does there may be a hazard of stray current leaking into the environment and inflicting harm or failing to provide electricity to the appliance.

.

3.FLOW CHART



5.OPERATION

For this Python project, we will use the RAVDESS dataset. This is the Ryerson Audio-Visual database of Emotional speeches and songs. Librosa is a Python library for scrutinizing audio and music. It has a flatter package layout, standardizes interfaces and names, backward compatibility, modular functions, and readable code. Using this, we calculated all the parameters of that particular wave. By using random forest and convolutional neural networks, we have classified the parameters to its appropriate emotion. Using firebase, the emotion is sent to the esp via IoT and according to our code, the respective LEDs glow to show output.

6.CONCLUSION

In this model, we have used CNN to extract features from the Ravdess dataset in order to classify the underlying emotions(neutral, calm, happy, sad, angry, fearful, disgust, surprised) in the given speech sample. This model has found many applications in the field of security, medicine, entertainment, and education. In the proposed model, librosa library has been used to convert an mp3 file to a wave file which is in binary format, hence readable by the computer. This helps in emotion recognition using speech. In order to carry out the result firstly we load data in the model, then extract features from it, and at last by using a classifier we are able to tell the accuracy of the model which is 95% in our case security, medicine, entertainment, and education. In the proposed model, librosa library has been used to convert an mp3 file to a wave file which is in binary format, hence readable by the computer. This helps in emotion recognition using speech. In order to carry out the result firstly we load data in the model, then extract features from it, and at last by using a classifier we are able to tell the accuracy of the model which is 95% in our case.

7.ACKNOWLEDGEMENT

We express our deep and sincere gratitude to our research guide, Prof. Prajakta Naregalkar for giving us the opportunity to work on this wonderful project ‘Emotion Recognition using Speech’ and for helping and guiding us throughout the time period. We could not have imagined having a better guide for our project. At last, we would also like to thank the Electronics Dept of our college for providing us the platform to carry out the accurate results for this project.

8.REFERENCES

[1] Juan Pablo Arias, Carlos Busso, Nestor Becerra Yom. Shape-based modeling of the fundamental frequency contour for emotion detection in speech[J]. Computer Speech & Language, 2014, 28(1):278-294.

[2] Yongming Huang, Guobao Zhang, Yue Li, et.al. Improved Emotion Recognition with Novel Task-Oriented Wavelet Packet Features [J]. Intelligent Computing Theory, 2014, 8588:706-714.

[3] H.M.Teager Sc.D, S.M.Teager. Allen. Evidence for nonlinear production mechanisms in the vocal tract [J]. Speech Production and Speech Modelling, 1990, 55:241-261.

[4] Y. Le Cun, B. Boser et al., “Handwritten digit recognition with a back-propagation network,” in Advances in neural information processing systems, 1990.

[5] O. Abdel-Hamid, A.-r. Mohamed et al., “Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition,” in 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP). IEEE, 2012, pp. 4277–4280.

[6] T. N. Sainath, A.-r. Mohamed et al., “Deep convolutional neural networks for lvcsr,” in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013, pp. 8614–8618.

[7] T. N. Sainath, B. Kingsbury et al., “Deep convolutional neural networks for large-scale speech tasks,” Neural Networks, vol. 64, pp. 39–48, 2015.

[8] Mohammed E. H., Courgeon M., Martin J.C., Mutlu B., Picard R, MACH: My Automated Conversation coacH, in Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pp. 697-706, 2013

[9] Healey, J. and R.W. Picard,”Detecting Stress During Real-World Driving Tasks Using Physiological Sensors,” IEEE Trans. on Intelligent Transportation Systems, Volume 6, No. 2, pp. 156166, June 2005.

[10] Global Personal Robots Market Size, Share, Development, Growth and Demand Forecast to 2022

Industry Insights by Type (Cleaning Robot, Entertainment & Toy Robot, Education Robot, Handicap Assistance Robot, Companion Robot, Personal Transportation Robot, Security Robot, and Others) published by P&S Market Research, Feb 2017. https://www.psmarketresearch.com/marketanalysis/personal-robot-market.

[11] Wiseguy Reports, Global Pet Companion Robots Market Outlook 2024: Global Opportunity and Demand Analysis, Market Forecast, 2016-2024, 12 September, 2017.

[12] Zuzanna Wojcik, Robotics&AI, May 2016.

[13] Garay, Nestor; Idoia Cearreta; Juan Miguel Lpez; Inmaculada Fajardo ”Assistive Technology and Affective Mediation”. Human Technology. 2 (1): 5583, April 2006.