JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Speech recognition and classification Using Deep Learning

Rishu Kumar[1], S.A. Badarkhe[6], Sandeep Singh[2] ,Tejas goje[3],V.P. Niwane[4] ,S.R Patil[6]

Department of E&TC Engineering SKNCOE, SPPU,Pune,Maharashtra

[1]singhrishu840@gmail.com
[2]sandeepanderson64@gmail.com
[3]gojetejas26@gmail.com
[4]vivek.niwane_sknco e@sinhgad.edu
[5]gk.sujata@gmail.com
[6]smita.garde_sknco e@sinhgad.edu

ABSTRACT

If we want to achieve emotional-related response from some algorithm or other intelligent machines, the initial step is fetch precise emotion recognition. This project deals with the implementation with the deep learning model of Convolutional Neural Networks (CNN). The architecture which is primarily based on an image processing CNN, developed in Python using Keras api which is based on TensorFlow platform.

The basic methods that lays the foundation for the classification of emotions recognition based on certain voice parameters is briefly described. As per obtained results, the model tries to obtain the average precision of 79.33% for five emotions namely (happy, fear, sad, neutral, anger), which is comparable with performances reported in scientific literature.

I.INTRODUCTION

In today's digital landscape, speech has become a primary mode of communication between humans and computers which has been possible by several technological upgradation. Speech recognition techniques along with signal processing made exponential progress in Speech-to-Text (STT) technology which is used nowadays in most mobile phones . Speech Recognition is the fastest growing research domains in which attempts are being made to recognize and decipher speech signals. This in turn leads to Speech Emotion Recognition (SER) growing research topic in which several advancements can lead to progress in numerous fields such as automatic translation systems, machine to human interface, used in synthesizing speech from text and vice versa and so on. Contrary to that this project focus is to survey and review various speech extraction features, emotional speech databases, classifier algorithms and so on. Problems present in various topics have been addressed.Speech Recognition is the terminology which deals with various algorithms and computing process to recognize the speech from the speech signals. Several technological strides in the field of the AI and signal processing, recognition of emotion made easier and possible.

II.MOTIVATION

Speech emotions basically refers to as extracting the emotional factors of the speaker from their speech.It approach is to study the speech of signal to detect the suitable emotions based on its characteristics like tone, pitch, etc. To extract these nuances and test the speech signal, a decent amount of algorithm have been designed with upgradation in latest technologies.CNN is one of the most used deep learning models that have resulted in massive success in research areas like 15 object recognition, face detection, and natural language pre-processing. Usually CNN has 3 fundamental building blocks,

convolutional layer, pooling layer, and fully connected layer. Thus, we depict these building blocks with some rudimentary concepts like softmax unit, rectified linear unit, and dropout.

III.SYSTEM REQUIREMENTS

For methodical use, every system needs configured hardware components along with respective drivers and dependency software systems. These prerequisites and requirements are popularly termed System requirements, these proposed requirements act as a regulative structure for software. In many cases two different types of requirements are mentioned which are: Recommended and Minimal. Because of advancement in technology and industry norms these technical specifications continuous changes and increase over a period of time. We can also define system requirements as platform specifications which must be met in order to run the given software smoothly without any technical obstacles and difficulties.

IV.SOFTWARE REQUIREMENTS

In accordance with process of software engineering software requirements are specifications which are required to run program on system or dependencies which must be satisfied in order to execute software successfully. Summary of such requirements is as follows:

1. A prerequisite or potential that is preoccupied by software or its component to met a standard specification or any other related criteria.
2. A prerequisite or a potential that is preoccupied by the system to complete the given task and finalize the outcome of intended objectives.
3. A well-documented depiction of a prerequisite or potential as in 1 or2.

Software requirements can be assembled as follows:

Platform	Windows 8+ / Linux 16.04+
Drivers	Display updated drivers
Dependencies	Python 3.7, Django, Html CSS, MySQL
Libraries	NumPy, pandas, scikit-learn, Google, ML kit
ML Models	Multiple Linear Regression, k-nearest neighbors

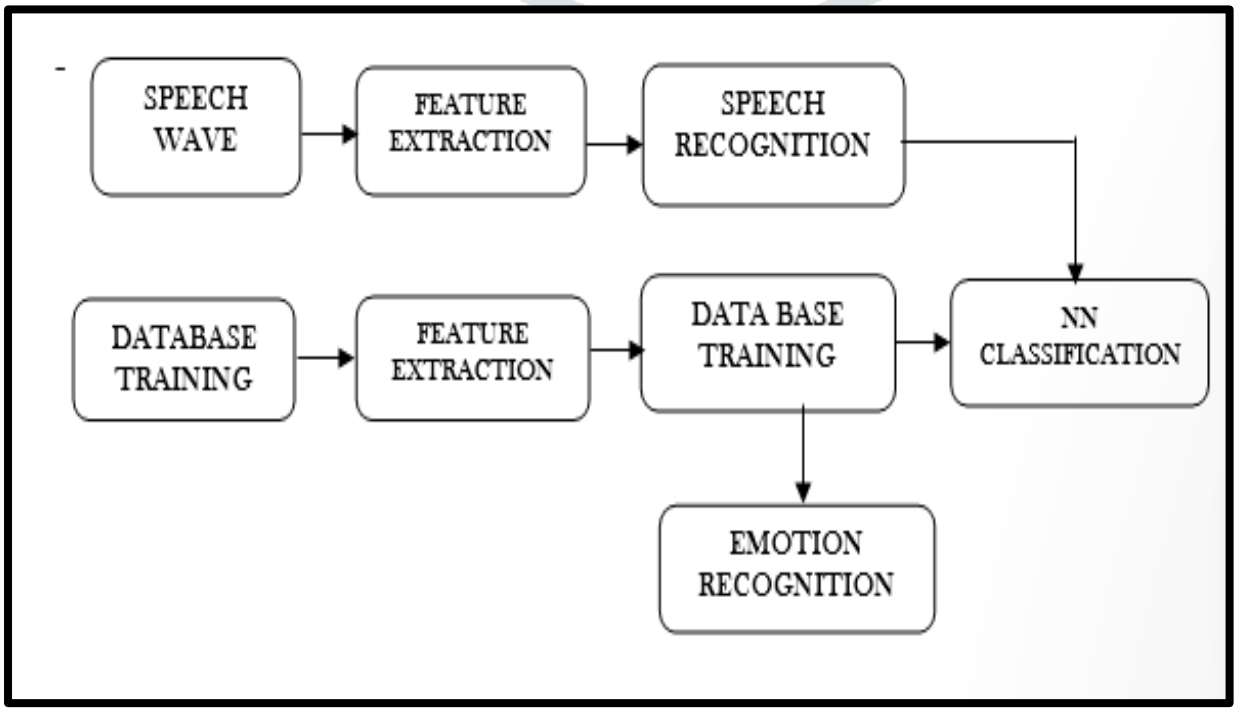

Fig.3.1: Block Diagram of System

Although activity apprehension from accent may be an analogously new acreage of analysis, it's several abeyant applications. In human-computer or human-human alternation systems, activity acceptance systems ability accord users with bigger casework by getting adjustive to their emotions.

In virtual worlds, feeling recognition might facilitate simulate additional realistic avatar interaction. The body of labor on police work feeling in speech is sort of restricted. Currently, researchers area unit still debating what options influence the popularity of feeling in speech.

V.CONCLUSIONS AND FUTURE WORK

Machine learning has made great progress so far, but in the field of speech signal processing, especially for building the SER system, there has not been much progress, SER is still a challenging problem. This paper proposes a new data enhancement method for SER and uses the proposed model that consists of the CNN, the LSTM, and the Attention Mechanism to classify speech emotions without using any traditional hand-crafted features.

REFERENCES

[1] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof and M. A. Mahjoub, "A review on speech emotion recognition: Case of pedagogical interaction in classroom," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, 2017, pp. 1-7, doi: 10.1109/ATSIP.2017.8075575.

[2] Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3-4), 169–200. https://doi.org/10.1080/02699939208411068

[3] Matilda's. Emotion recognition: A survey. International Journal of Advanced Computer Research. 2015;3(1):14- 19

[4] Koolagudi SG, Rao KS. Emotion recognition from speech: A review. International Journal of Speech Technology. 2012;15(2):99-117

[5] Ali H, Hariharan M, Yaacob S, Adom AH. Facial emotion recognition using empirical mode decomposition. Expert Systems with Applications. 2015.

[6] Liu ZT, Wu M, Cao WH, Mao JW, Xu JP, Tan GZ. Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing. 2018.