ITSP Final Documentation
** This will be considered as your final documentation and will be made available for online usage.*

Speech Emotion Recognition

ITS20061

Keywords (Include 7 or more keywords which will help others find your documentation easily)
Emotion Recognition, ANN, Librosa, Features, Predict, keywords 6, keyword7

Team Member Name	Roll Number	Email-Id
Shivam Modi	190070059	modishivu09@gmail.com
Shiv Modi	19D100011	modishiv210@gmail.com
Prasada Thombare	19D100014	prasadathombare1842001@gmail.com
Sakshi Kasralikar	190070030	sakshiswami1303@gmail.com

Inspiration for Idea

Speech Emotion Recognition is one of the basic and essential applications of Machine Learning. The importance of emotion recognition of human speech has increased in recent days to improve both the naturalness and efficiency of human - machine interactions. This is one of the well-known applications of Deep Learning(a substitute of Machine Learning). The wide areas of applications of SER made us choose this topic for understanding Machine Learning.

Problem Statement

Emotions play an important role in human mental life. It is a medium of expression of one's perspective or one's mental state to others.
Here are some of the applications of SER that can be used in our daily lives:-

1. It can serve as the performance parameter for conversational analysis, thus identifying customer satisfaction.

2. It can be used in-car board systems based on information of the mental state of the driver that can be provided to the system to initiate his/her safety preventing accidents from happening.

3. Many professions like Teaching require deciding and developing strategies for managing emotions within the learning environment.

Data is the lifeblood of all business. Data-driven decisions increasingly make the difference between keeping up with competition or falling further behind. Machine learning can be the key to unlocking the value of corporate and customer data and enacting decisions that keep a company ahead of the competition.

Existing solutions in the Market

Speech emotion recognition has also been used in call center applications and mobile communication. The main objective of employing speech emotion recognition is to adapt the system response upon detecting frustration or annoyance in the speaker's voice. The task of speech emotion recognition is very challenging for the following reasons.

Proposed Solution

So We decided to create our own model for Speech Emotion Recognition which we can use to understand emotions well and create a website where we people can just play with things or use this anywhere and anytime for their convenience.
Earlier we had decided to add two different components which are speech as well as words for detecting the emotions but later we stick to only recognizing emotion from speech.

Brief Description

We have created a website where we can input the audio file whose emotion we have to detect and then in return we will get the output which is emotion associated with the audio file.

Currently our goal is to find 4 emotions only: Happy, Sad, Angry and Neutral.We got 195 features after applying feature scaling and we also to_categorical the emotions for our ANN.

We used librosa and soundfile to extract emotion from the audio file and trained & tested with an ANN model we created, test_size was 0.25 and got the **accuracy of 70%**.

Progress

Describe how the work was done
Work distribution in the team
Work-Flow distributed across the duration between the review meets
Challenges (Difficulties faced and how you overcame it)
Calculations involved
(min.300 words)

Speech Emotion Recognition is a project that aims to detect emotion associated with a particular input audio file. Earlier our plan was to achieve it by two ways: 1) Sound Features 2) Words/Text Features. But the Text thing didn’t work out so we then just went with Sound Features. Before the first review meeting, we just worked on learning but after being suggested by the review panel, after that we worked on work distribution. We distributed our work according to this:

Prasada Thombare: 3 crucial steps in work: data preprocessing, dataset loading and dataset splitting.

Shiv Modi: work on Text embedding, increasing accuracy along with Shivam.

Sakshi Kasralikar: build a website and save our model so that we can present our project more efficiently.

Shivam Modi: creating the model and setting up the appropriate hyperparameters for the model, increasing accuracy for the model.

By the second review meeting, our almost model was ready, just had to increase the accuracy. We had accuracy of 55% then which is so low and unreliable. Sakshi had made a decent website but there was a lot left like a saving model so that the emotions can be detected using the website.

Challenges:

There were a lot of challenges, some were sorted but some couldn't so then the idea had to be dropped.

First challenge was choosing an appropriate dataset. We chose RAVDESS prior to the project starting but we came to know from our mentor that it wouldn't work so we need to change it to CREMA-D after almost 2 days of research.

Another major Challenge we faced was about our fundamental plan. As said earlier we were using text as well as sound features together as emotion detection but the text thing didn't work out and due to less time before the review meeting, we had to drop that idea and only focus on sound features.

Then came the issue of accuracy so we had to increase it somehow because it's a really important thing in such tasks like SER. We tried many ways to do that, some worked out while some didn't.

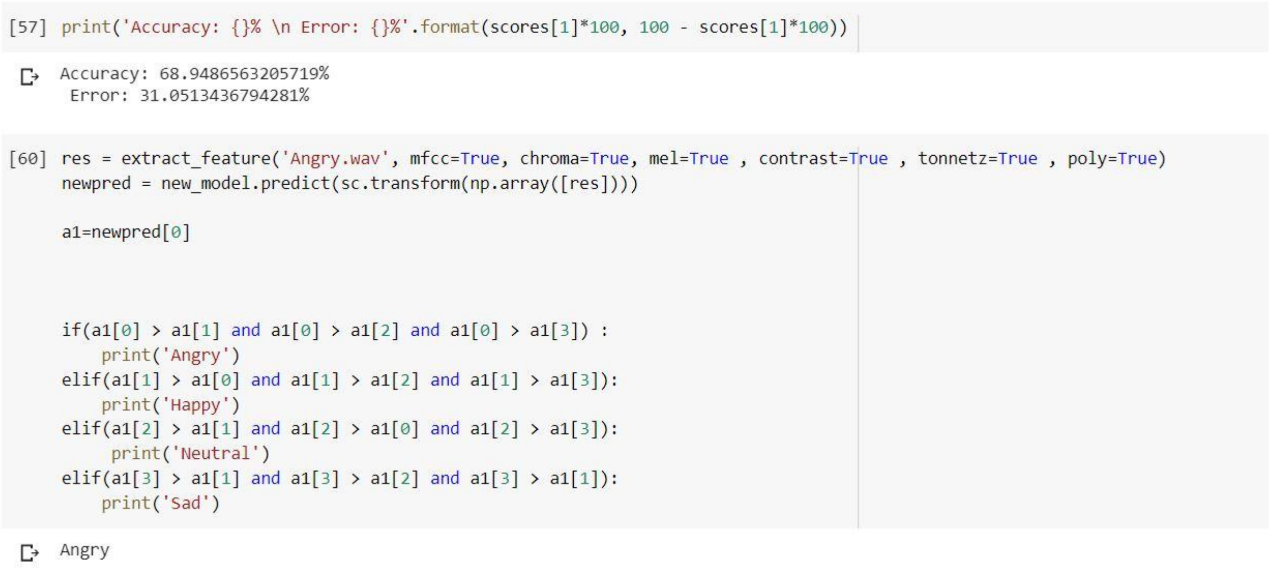
First of all we change hyperparameters which worked out, then add feature scaling which also increased the accuracy. But we tried out feature engineering and models ensemble but that didn't workout.

The major challenge was faced at the last moment, we have to write a local python script for Backend where we just have to save a model, give an audio file and run that file on that model and get the result. Then this will go to nodej, where the local server will be spinned up and the file will be uploaded. We will give commands in node to run the python script but the script is predicting incorrectly on Spyder.

There weren't major calculations but here the mathematics behind this can be explained like:

We used librosa and soundfile to extract the sound features which came out to be 195, trained in our ANN with appropriate hyperparameters that can give us maximum accuracy with test_size=0.25.

Results



```
[57] print('Accuracy: {}% \n Error: {}%'.format(scores[1]*100, 100 - scores[1]*100))

     Accuracy: 68.9486563205719%
      Error: 31.0513436794281%

[60] res = extract_feature('Angry.wav', mfcc=True, chroma=True, mel=True , contrast=True , tonnetz=True , poly=True)
     newpred = new_model.predict(sc.transform(np.array([res])))

     a1=newpred[0]



     if(a1[0] > a1[1] and a1[0] > a1[2] and a1[0] > a1[3]) :
         print('Angry')
     elif(a1[1] > a1[0] and a1[1] > a1[2] and a1[1] > a1[3]):
         print('Happy')
     elif(a1[2] > a1[1] and a1[2] > a1[0] and a1[2] > a1[3]):
          print('Neutral')
     elif(a1[3] > a1[1] and a1[3] > a1[2] and a1[3] > a1[1]):
         print('Sad')

     Angry
```

CODE LINK:
https://colab.research.google.com/drive/1IbB3neVqwdHEFZmyRvMBs0gSEDXSSw1_?usp=sharing#scrollTo=r0dUpgSnT3eE
PROJECT VIDEO LINK:
https://drive.google.com/file/d/1JL8r7DplUwReSUA0aef3xh5dt8W58t6V/view?usp=sharing
WEBSITE LINK:
https://sakshikasralikar.github.io/speech_emotion/
FINAL PRESENTATION LINK:
https://docs.google.com/presentation/d/162ONVfmsfQ7XzxnOukiluhGgiQ2taEivhBXValyWuEk/edit#slide=id.g8a8c9be888_1_5

Learning Value

The biggest motivation for doing this project was to learn Machine Learning and this project was proven to be appropriate for us as we didn't lack motivation till the end for our project and after this, we also aim to further learn and work more on various other ML projects.

As Machine Learning is also a Python project, we learnt how we can work and implement Python to get our desired Project done.

We learnt about various Python and ML libraries like numpy, librosa, soundfile and sklearn which helped us a lot while working on our model and are expected to be useful for future projects also.

Sakshi Kasralikar - One of the members of our team, began to learn web development and created an amazing website in just a short period of time. Learnt HTML , CSS , JS for front- end and nodejs, expressjs for backend.

Software/ Hardware used

Softwares that we are using in our project:

1) Google colab
2) Anaconda Jupyter Notebook
3) Kaggle

1) We used google colab to run example datasets mentioned in the work done till date.

2) We used an anaconda jupyter notebook to run the RAVDESS dataset using an MLP classifier for practise purposes.

3) We used Kaggle to download required datasets like RAVDESS & Crema-D for our Project.

Suggestions for others

Emotions play a very huge role in our life and Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch. Therefore It becomes really important to recognize emotions well and thus the accuracy of the model is a very essential part of our project. So the first important suggestion is that one should attempt to bring accuracy of the model to its Best.
Now, the basic suggestions which are more based on ML are that one should constantly try to search and learn about various libraries that are useful for their projects like in our project, we searched about various libraries that can be useful in speech and emotion fields and we found very useful libraries like **Librosa** and **Soundfile** which reduced a lot of our efforts and time.

Contribution by each Team Member

Prasada Thombare:

- He did significant work in data preprocessing, dataset loading and dataset splitting. All these 3 crucial steps were done by him so that we could further proceed in our project.
- He then joined Sakshi to help her in the website part i.e he helped in writing python script for making predictions which was to be used by Sakshi in the back-end part.

Shiv Modi:

- Earlier he worked a lot on Text embedding but that didn’t work out for our model. As mentioned earlier also that our previous idea contained a text/words part also for detecting emotion along with sound features.
- He then worked on increasing accuracy along with Shivam.

Sakshi Kasralikar:

- Her main task was to build a website and save our model so that we can present our project more efficiently.

Shivam Modi:

- He worked on creating the model and set up the appropriate hyperparameters for the model.
- He worked on increasing accuracy for the model. Earlier model's accuracy was just 55% but after working on it by trying various methods like Feature Scaling, it was brought up to 70%.

References and Citations

CITE ALL THE RESEARCH PAPERS AND AUTHORS (Whatever is demanded by the paper) and also add the link to it
Mention all GitHub repositories used for reference. If you have used code from other github repositories make sure you mention and cite the authors properly and INCLUDE THE GITHUB LICENSE in the disclaimer section.

Some of the sites that we referred:

Where we took the tutorials for Deep Learning:
Deep Learning - Grundy

Where we understood about librosa and soundfile:
https://www.thepythoncode.com/article/building-a-speech-emotion-recognizer-using-sklearn

Where we clear out even our smallest doubts and learnt to increase accuracy:
https://machinelearningmastery.com/

Where we looked out for models and layer:
https://keras.io/api/

Where we run our code:
https://colab.research.google.com/drive/1IbB3neVqwdHEFZmyRvMBs0gSEDXSSw1_?usp=sharing

Where we found and downloaded the dataset:
https://www.kaggle.com

Disclaimer

Fair Use Disclaimer: Include "Fair use of … " wherever necessary (instances where you have directly used materials or taken words from other sources)
Copyright Disclaimer: Make sure it is well within limits of use allowed in their respective copyright policies.

We referred to the data loading and data preprocessing code from this site because this site permits for such actions:
https://www.thepythoncode.com/article/building-a-speech-emotion-recognizer-using-sklearn

We just took basic ideas from here, else we did everything on our own.

Referred https://www.w3schools.com for web development

Licenses

Include licenses of all the software and tools (if it is free to use, open-source, proprietary, paid) used and mention GitHub licenses