Accent Recognition

Anmol Arora(130050027), Karan Vaidya(130050019)

Problem Description and Motivation

- To take the recording of the user speaking a reference language(in our case English) as input and recognize the accent and therefore nationality of the user using the accent information.
- All the voice-based assistants and other voice services are trying to predict accent of the person
- Helps in improving the Speech Recognition, as we can use accent specific information and train and test specifically
- Also gives additional information about the user as accent can be used to predict ethnicity

Methodology

Reference: removing word related features

- Choose a speech reference: synthesized output of a text-to-speech engine, devoid of human biases and noise

- Calculated mfcc features for the most common words in english

- This given accent free audio signals representing the word-particular audio features.

Preprocesing

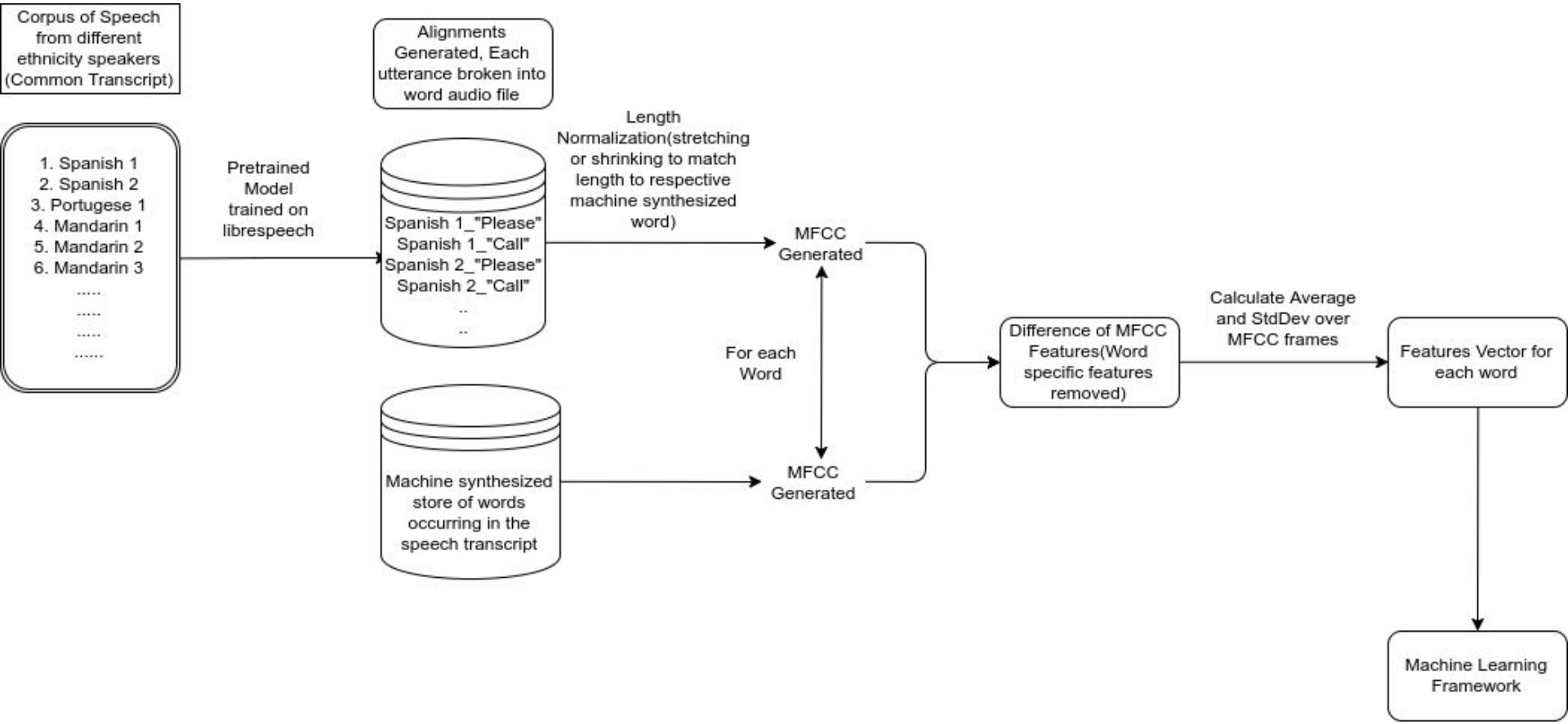
- DataSet - GMU Speech Accent Archive
- Sniped the audio files to segregate the words
- Used acoustic model, pretrained on ‘Librispeech’ corpus
- Audio files of the word stretched/compressed to match the length of the reference word, to get same length of MFCC
- Subtract the MFCC features of the given word with reference audio of that word
- Take the mean and standard deviation of MFCC features over the whole window of word(13 + 13 =26 features) to get the sufficient statistics of the word

Main Hypothesis

- Obtained the same length feature vectors of the word, irrespective of the length
- The sufficient statistics will capture the accent information and will correlate more with the words of the similar accent, rather than similar sounding words (since we subtracted the features of reference word)
- Tried various supervised models and got ~30% accuracy in predicting over 193 equi-probable accents,against 0.5% of randomly guesses, which justifies our hypothesis
- Trained the proposed model on various supervised models like KNN, Random Forests, GBM, SVM

Predicting the accent from test Speech

- Similar preprocessing as training data, segregating words from speech treating every word independent
- Apply the trained ML model to get the probability distribution of accent for each word
- Take a weighted average over the probabilities(confidence score) with weights as length of the composing words to finally label the utterance


Corpus of Speech from different ethnicity speakers (Common Transcript)
1. Spanish 1
2. Spanish 2
3. Portugese 1
4. Mandarin 1
5. Mandarin 2
6. Mandarin 3
.....
.....
.....
.....
Pretrained Model trained on librespeech
Alignments Generated, Each utterance broken into word audio file
Spanish 1_"Please"
Spanish 1_"Call"
Spanish 2_"Please"
Spanish 2_"Call"
..
..
Length Normalization(stretching or shrinking to match length to respective machine synthesized word)
Machine synthesized store of words occurring in the speech transcript
MFCC Generated
For each Word
MFCC Generated
Difference of MFCC Features(Word specific features removed)
Calculate Average and StdDev over MFCC frames
Features Vector for each word
Machine Learning Framework

Experimental Setup

- Scraped GMU Accent Archive and collected over 2300 utterance from people of 193 ethnicities
- Audio files were processed into kaldi compatible wav format(44.1KHz)
- Each of these utterance was aligned with the transcript and snipped into words
- The dataset so obtained was split into training(70%) and testing (30%)

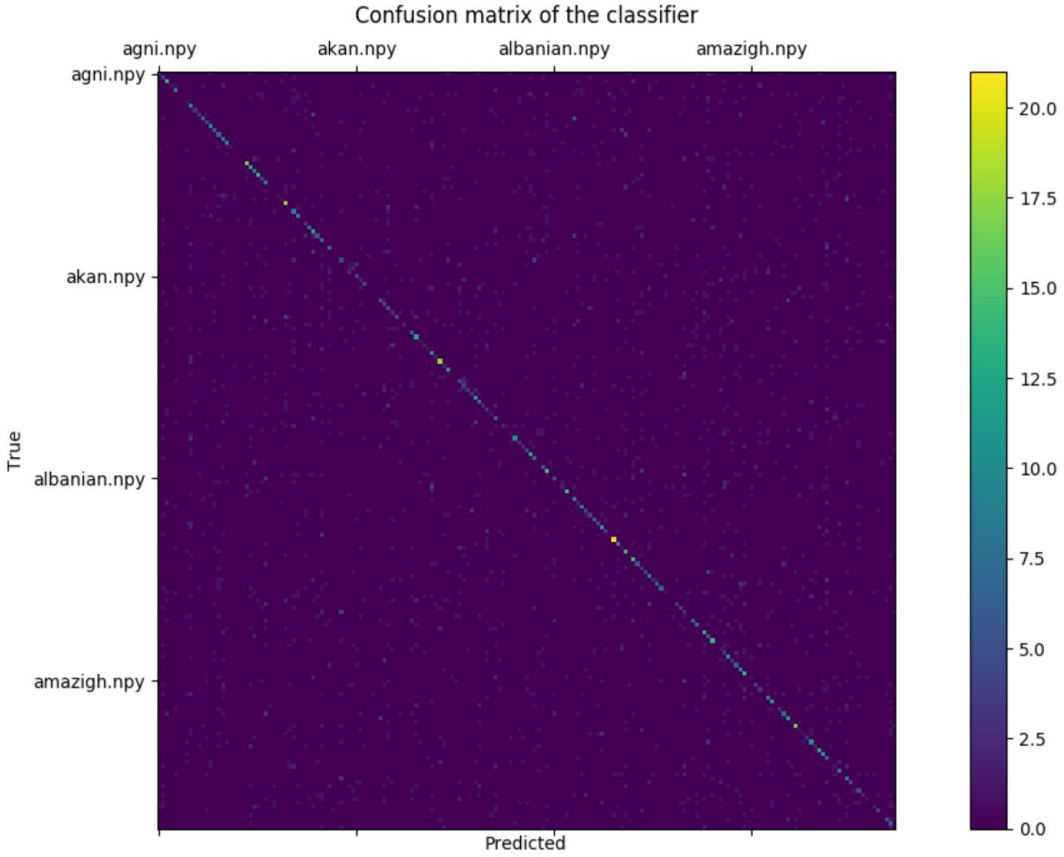
Results

Accuracy of Various Models

- KNN - 14.6 %
- Decision Tree - 8.9%
- Random Forest - 32.8%
- MLP Classifier - 31.1%
- SVM - 26.6%

The best result obtained is from Random Forest (n_estimators=1000,max_depth=30) which gives an accuracy of over 32.8% over test data (against 0.5% from randomly guessing over 193 classes). This shows that the proposed methodology can be used to extract accent related information from voice signals.

Figure 2


Confusion matrix of the classifier
agni.npy
akan.npy
albanian.npy
amazigh.npy
True
Predicted
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0

Thank You