Team 29
TA Mentor: Avinash
Projit Bandyopadhyay (20161014)
Nitin John Raj
Aniket Bansal (20161152)
Ujjwal Tiwari (2018900100)

Audio Based Language Detection (37)

Github Link: https://github.com/ProjitB/Audio_Based_Language_Detection

OVERVIEW

Aim of the project is that given, a segment of audio in English, we should be able to detect the mother tongue of the speaker.

Main Goals

1. Create a dataset matching audio segments in English to mother tongue's of the speakers.
2. Identify usage of words / pronunciations particular to those of a certain mother tongue.
3. Train a classifier to be able to identify mother tongue from a speech sample (where the sentences to be spoken are predetermined)
4. Generalize the previous classifier for unstructured speech.
5. To be able to create a live demo wherein a speech sample is given, and the speaker's mother tongue comes as an output. (Possibly with a confidence score)

Specifications

- Input will be: arbitrary length of audio track.
- Output will be: Mother Tongue, from pre-decided list of potential languages + confidence score.
- Processing time for the whole procedure must be low (almost real time, as live demo is expected)

Problem Definition

Data:

We plan to start out by collecting a large corpus to play around with. This would probably entail scraping multiple Youtube videos and/or news broadcasts. A potential way to go about this is to scrape wikipedia for with a list of celebrities, and find their corresponding mother tongue(usually found on their main wikipedia page). Next we would have to scrape youtube for public interviews and other forms of audio. These videos then need to be segmented to only hold the audio of that single speaker.

Issues we may face here include: heavy bias towards certain languages ex. Hindi, finding pure segments of audio, ability of people to mask their mother tongue (eg. singers)

Classifier:

Building a classifier is not possible without the previous stage. Here it may be prudent for us to build a simple classifier first before moving on to complex neural nets..etc. There has been a good amount of research done in identifying mother tongue from speech on the basis of getting people to pronounce certain sentences. This will probably be our first approach, as this can be done in tandem with dataset collection. The next job will be to generalize features from the sentences used to determine mother tongue, and see if it can be identified in continuous speech which is not forced. We can split up at this point and have to approaches: a naive one where we tell the speaker what sentences to use, and a more complex one where the speaker may choose to say anything.

Optimization:

Lastly, we will need to optimize our classifiers for the amount of data, and constrained time, to be able to perform near real-time. WIth limited data, our classifier should be able to output a probability distribution over the mother tongue possibilities. Additionally, multiple types of classifiers can be used, and a certain amount of hyperparameter training will also need to be done. This will come in at this step.

MILESTONES / RESULTS

Dataset Collection + Annotation (30 March)

This project doesn't come with a pre-made dataset that can be used directly to train a classifier. Thus we must start out collecting a dataset, largely comprising of speech samples of public figures(as their mother tongues would be known), and the corresponding mother tongues.

Additionally, there may be certain features of each audio track that we would want to annotate: ex. Presence of certain words, corresponding text of the audio track, location of the place where the audio track is recorded.etc.

Decision Tree with Forced Sentences (30 March)

Given a speaker, and a set of sentences for them to say, we want to be able to identify the mother tongue of the person. These sentences will be crafted with linguistic knowledge so as to force the appearance of certain pronunciations and words which may not be present in common speech. By traversing this decision tree, we should be able to identify, with certain confidence, the mother tongue of the speaker.

Classifier with Unstructured Speech (20 April)

This will be the hardest part of the project, where we must generalize the features which we have extracted with linguistic knowledge and search for these features, or similar, within unstructured (unforced) speech. Both this and the above will be part of our demo.

Task Split Up (Parts of the project where each teammate will contribute):

- Projit: Scraping Wikipedia, DT with structured sentences, final classifier
- Nitin: Finding speeches and other audio samples, decision tree for structured sentences, generalized classifier
- Aniket: Creation or finding suitable dataset, DT with structured sentences, classifier
- Ujjwal: I will be involved into literature survey and look into various design aspects of the Classifier. From the looks of the problem Probabilistic Graphical Models(PGMs) - Deep Belief Networks, Boltzmann machine, Variational autoencoders seem a good starting point. The following github link can also be helpful : https://github.com/tensorflow/tensor2tensor

  I will also be working on the documentation and presentation of the project.