Pinning Accents: A Study on Accent Classifiers

Akanksha, Akhilesh, Connor, Kathan, Sheryl
October 1st, 2019

Introduction: Due to the recent advancements in the area of speech recognition systems, it has become necessary to recognize the nuances posed by various accents in spoken language. AI-enabled ASR assistants such as Siri and Alexa are unintentionally equipped to perform better with American/British accents, despite being increasingly prevalent in a diverse society. One of the key challenges in automating speech recognition is understanding speech by non-native speakers. Accents can signify a speaker's ethnic identity, regardless of the context or language being spoken. This project is dedicated to building and evaluating reliable machine learning models that classify accents by identifying the ethnicity of a speaker. Such a classifier has a plethora of real-world applications, allowing for more robust accent-specific speech recognition systems.

Methodology: The project would be essentially done in two parts, i.e, feature extraction from .mpg/.wav files from the Speech Accent Archive [1] and modeling. Preprocessing would be done by extracting the Mel-frequency cepstral coefficients (MFCC) from the dataset through Python tools, and modeling of the accent classifier would be done through both supervised (Naive Bayes, Support Vector Machines, Decision trees [4] and Deep Learning Techniques [3]) and unsupervised (K-Means and Gaussian Mixture Models [5]) techniques. We will potentially use Principal Component Analysis (PCA) to narrow down our feature set.

Potential Results: The primary goal of this work is to compare different supervised and unsupervised machine learning algorithms on the **accent classification task**. The aim is to make the model identify important features from sound waves that clearly differentiate various accents. We then aim to use these model characteristics to gain insights into what makes a **pair of accents similar/dissimilar**. The MFCC closely resembles how humans perceive and understand sound which is based on the shape of the vocal tract and gives a good estimate of the phoneme being produced. We hope to leverage this and provide insights at the **word** level and show how the same word differs when spoken in two different accents.

Discussions: Speech recognition algorithms perform better when the gender and accent of the speaker are determined in advance. That way, separate models can be trained for each gender-accent pair [2]. The referenced method, however, requires the training data of speakers to be manually categorized by gender and accent. This project could discover unsupervised learning methods of determining clusters of speakers based on accent and gender, which could be used before speech recognition to map a speaker to a likely cluster and then use an appropriately trained model for that cluster.

References

[1] Speech accent archive. Available at `http://accent.gmu.edu`.

[2] Too Chen, Chao Huang, Eric Chang, and Jingehan Wang. Automatic accent identification using gaussian mixture models. pages 343– 346, 01 2002.

[3] Vidya Prasad K, Akarsh S, R. Vinayakumar, and Soman K. P. A deep learning approach for similar languages, varieties and dialects. Available at `http://arxiv.org/abs/1901.00297`, 2019.

[4] An Sheng Leon Mak. Deep learning approach to accent classification. 2017.

[5] Morgan Bryant mrbryant. Classification of accents of english speakers by native language. 2014.