Gender, Accent, Age Recognizer from Audio

Anuj Khasgiwala, Bhagyashree Mandora, Meiling Long

1. Introduction

There has always been a need to identify various characteristics of a person just by listening to their voice. This happens in phone conversations or recorded audio from businesses. This application will help predict the traits of the person associated with an audio, with respect to the probable gender, age and the region of origin based on his accent. For which "Machine Learning" algorithms like the SVM and audio processing libraries will be employed to predict the outcomes of the voice.

Nowadays, there is an increasing focus on voice as an input for businesses. Virtual assistants which take voice commands now reside in mobile phones. Technology using voice can now be found in smart systems like cars, shopping complexes, security locks etc. Analyzing audio for such features can help give personalized results for the commands. This application will be very useful in data analytics to help make intelligent business decisions.

2. Resources Available

Following are the sources which may be helpful to complete the project:

2.1. Gender

- Kaggle- potential data set
- WavSource - potential data set
- Classical Convert - audio analysis
- Python speech recognition library - audio analysis

2.2. Accent

- GMU speech accent archive- potential data set
- Research paper - accent classification using scattering coefficients
- CMU Sphinx - audio analysis

2.3. Age

- Research paper - Age Recognition from voice, Western Reserve University, Cleveland, Ohio
- Research paper - A review of seaker age classification, IJSR
- Research paper - Comparison of methods for age and gender classification
- Research paper - Pitch range based feature set for age and gender classification.
- Research paper - A comparative study of gender and age classification
- Praat - audio analysis

3. Toolkit

- Python/R, Pandas, NumPy, SciPy, SciPy.io.wavfile, XGBoost, scikit-learn
- Highcharts, Tableau, HTML, JavaScript, CSS, matplotlib

4. Preliminary sketch

The application will take audio files as input and will make predictions as to the probable gender, accent and age of the person.
We hope to build the application in the following steps-

- Data collection - Training data for gender and accent information can be obtained through web crawling or scraping the above mentioned websites. The data for gender information is available in one data set but it may be necessary to merge it with the additional data found.
- Data processing - The data set files need to be cleaned and analyzed for the relevant acoustic properties needed for predicting the characteristics.
- Application of machine learning statistical models - Based on the research papers, we may need to relate the acoustic properties to gender, accent and age. This relation is used to train the machine using some of the classification and regression algorithms and make predictions on the test audio file used.
- Visualization - Various comparison and relationship charts will be made using visualization tools. Certain insights from the data will also be displayed.

Both, the prediction of properties given an audio file and the findings can be used in further data analysis by businesses.