


Recognize gender & accent of audio: A Study of Algorithms

Anuj Khasgiwala, Bhagyashree Mandora

CS 6600 Fall 2016

Abstract

Industry is moving towards IVR services and voice based smart systems like virtual assistants. Voice input analysis will contribute to make these systems more intelligent and help in formulating business policies. This project analyses the importance of different acoustic characteristics in prediction of gender. It employs various classification algorithms as a comparison study. It also investigates ways to analyze audio to predict the accent of the associated person.

Problem Description

Given a random audio file with human voice, predict the gender and accent of the speaker associated with the voice.

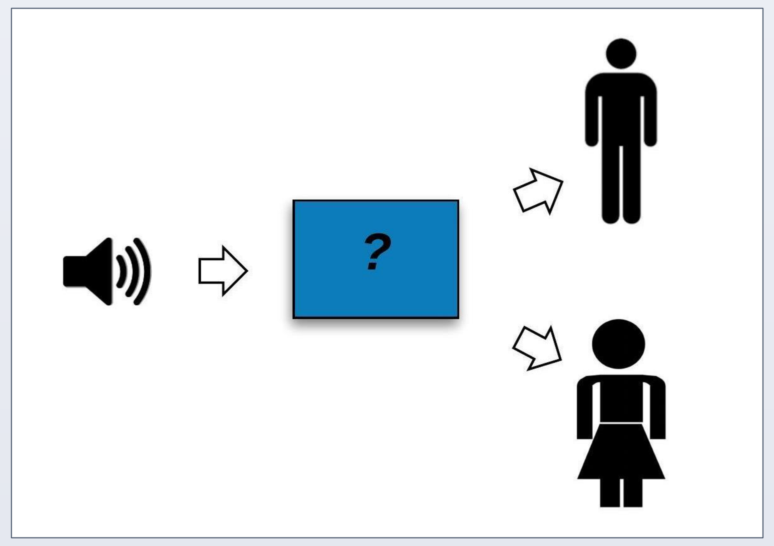

Fig 1: Problem description

Methods

For solving the problem user will input an audio file to the recognizer:

- Extract audio features from wav file to compute acoustic characteristics
- Find importance of acoustic characteristics in prediction
- Perform classification using different algorithms
- Measure accuracy of algorithm using 10-fold cross validation
- Extract scattering coefficient of audio
- Predict the accent of speaker

Analysis of Importance of Acoustic Characteristics:

A. Hypothesis Test- A proposition was tested with Welch Two Sample t-test and checked if *generally* a male voice feature is greater or lesser than the same female voice feature. The p-value obtained should be less than 5%.

B. Visual Representation through Boxplot

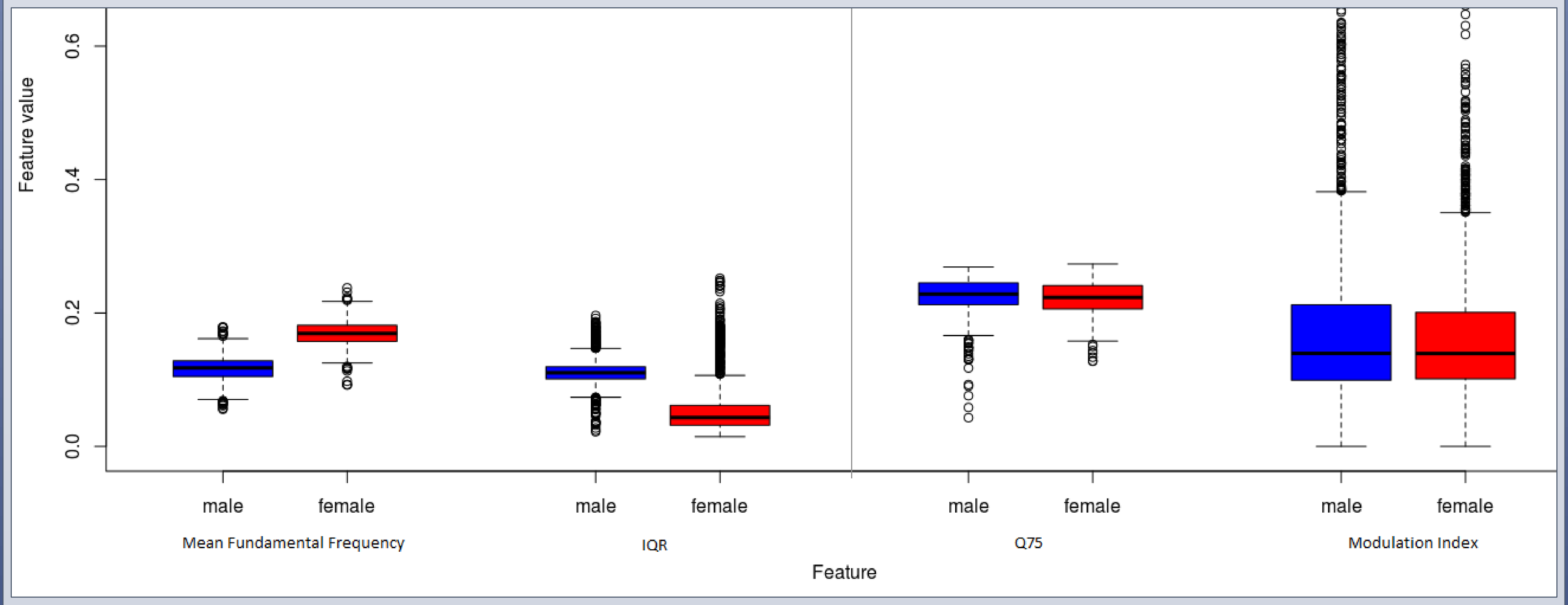

Fig 2: Boxplot to compare importance of different acoustic characteristic for predicting gender. The chart considers four acoustic features used for differentiating between male and female voice. Since there is a greater difference in the median values of "Mean Fundamental Frequency" for male and female voice, this feature may be a good factor in the classification. Similarly, a smaller difference in the median values of "Modulation Index" for male and female in the boxplot shows the lesser importance of the feature in gender prediction.

Results

A. The model can successfully predict the gender with a good accuracy. Gender classification based on voice characteristics is based primarily on Mean Fundamental Frequency and a value of 142 Hz forms a boundary between male and female voice.

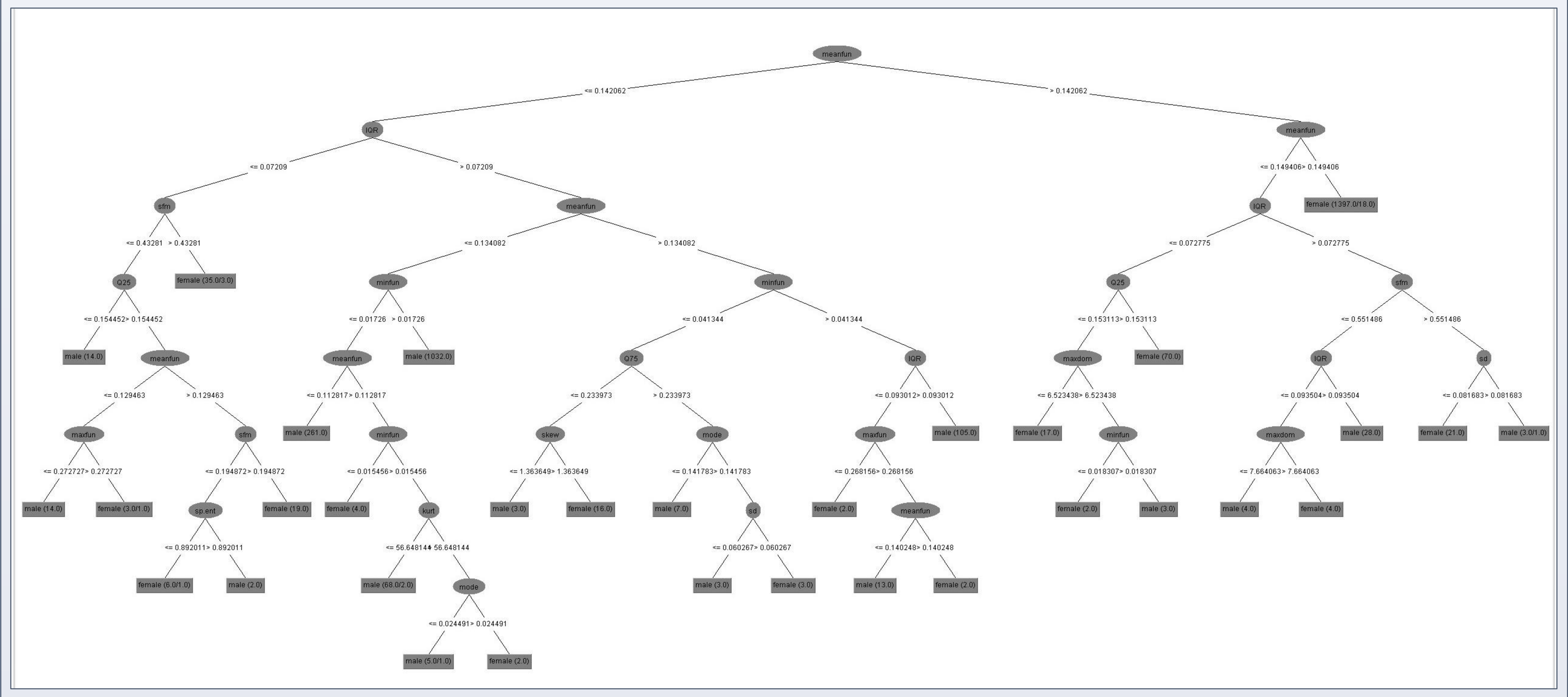

Fig 3: A tree showing the classification based on the features and the split value

B. Varying the Features used for Training and Prediction- Each voice feature was removed from the training and test set one by one and the accuracy was tested on 80% test split data using a Neural Network. A fall in accuracy was noticed when an important feature was removed. However, an increase in accuracy was also obtained when a less important feature was not considered. This may have been due to reduction in noise for prediction.

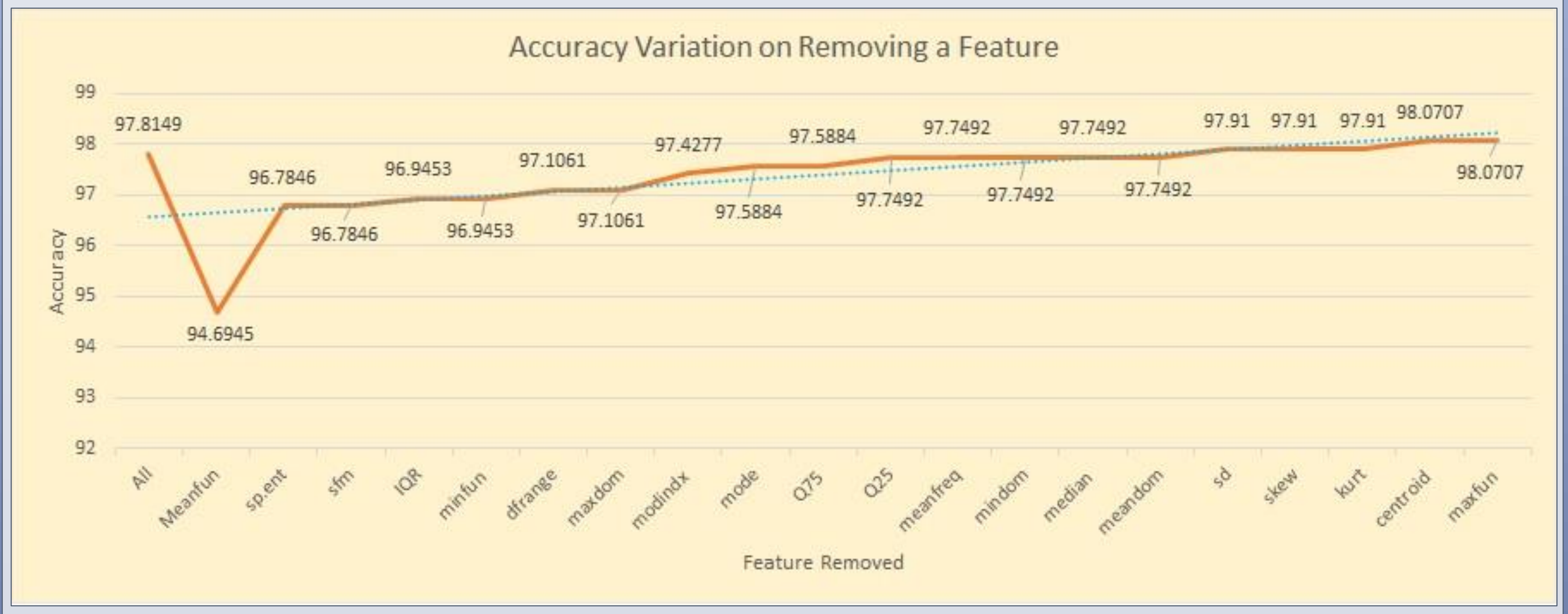

Fig 4: Impact on accuracy of Neural Network, by removing one acoustic characteristic at a time. The orange line in the Line chart is used to show the change in accuracy while the blue line is the best fit line (a generic single line to make up for all the points). The X-axis represents the feature removed where first value is accuracy using all features. The general trend shows a decrease in accuracy when an important feature is removed and an increase when a less important one is removed.

C. Varying the Algorithm used for Prediction-

Five algorithms were used to predict gender and the accuracy obtained is as shown in the chart. SVM showed a comparatively low accuracy. This may have been due to error introduced due to a large number of support vectors found near the separating hyper-plane between male and female values for the feature. Any noisy data leads to a wrong prediction.

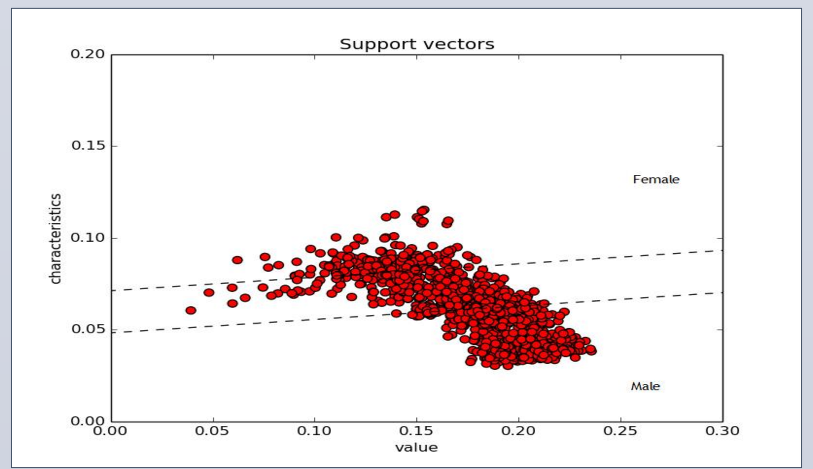

Fig 5: The red dots are the support vectors for the SVM model. Support vectors are the data points lying close to the separating hyper-plane between male and female voice.

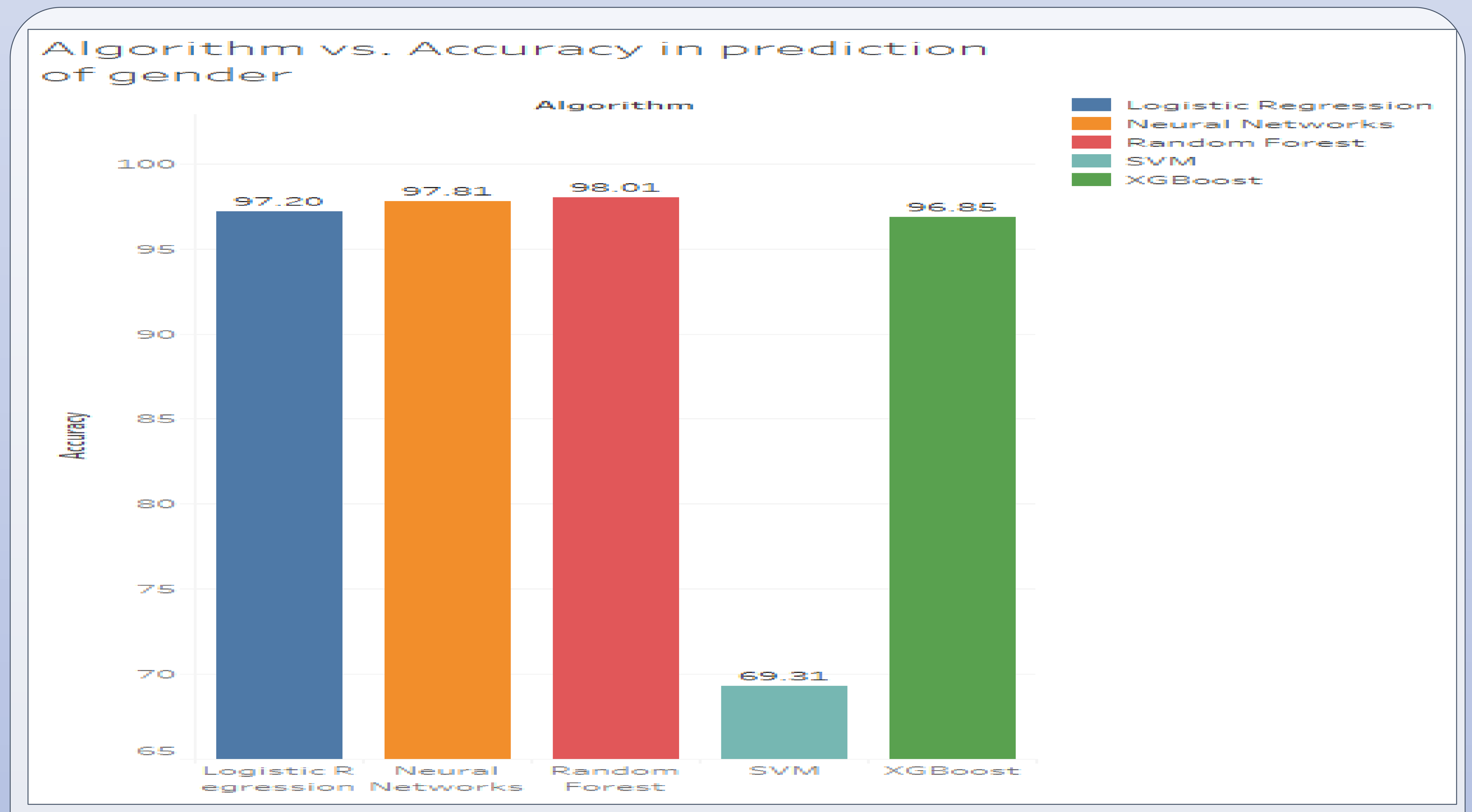

Fig 6: Variation in accuracy with algorithm used for prediction of gender. The results are obtained through 10 fold-cross-validation over the training set.

Conclusion and Future Work

The prediction of gender is possible using these 21 acoustic characteristics and the importance of each characteristic is also analyzed. The performance of various algorithms is also checked for such a data set.

Future Work includes-

- Finding the scattering coefficients of the audio file to predict accent-
- Scattering Coefficients are wavelet transforms with non-linear operators. For a given language, the scattering coefficients are considerably different for people from different area and different accents. This can be used to differentiate between accents from different audio.
- Training machine learning models using Scattering coefficients of audio and predicting accent
- Analyzing the results obtained as to how they contribute to accent prediction

References

[1] Kory Becker, http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/

[2] Theodoros Giannakopoulos, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144610

[3] George Mason University, http://accent.gmu.edu/browse_language.php

[4] Morganne Lerch, http://cnx.org/content/col11938/1.1/

[5] Kaggle, https://www.kaggle.com/primaryobjects/voicegender

[6] Buket D. Barkana, Jingcheng Zhou, https://www.researchgate.net/publication/275949466_A_new_pitch_range-based_feature_set_for_speaker_age_and_gender_classification

[7] Michael Norris, Michael Wagner , http://staff.estem-uc.edu.au/mwagner/files/2014/04/NorrisWagner_AgeGenderThroughPhoneRecognition_SST_2010.pdf

Acknowledgement

Thank you Dr. Nicholas Flann for guiding us throughout this project.