International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Vocal Data Assesment To Envision Distinctive Features of An Individual

Arnav Garg, Kushal Agrawal, P. Akilandeshwari

Abstract - There is a lot of audio data generated on a day to day bases, which goes to waste without undergoing due processing. If we process this data, it can be beneficial for a multitude of purposes. Vocal data is unstructured, which makes it even harder for processing. This data has to undergo thorough pre-processing to convert it to a machine-understandable form. We aim to perform analysis of human voice to extract meaningful data and make a prediction of their age, gender, and accent. The developed system uses the Mel-frequency Cepstral Coefficient (MFCC), zero-cross-rate(ZCR), chroma_stft, spectral_centroid, spectral_bandwidth, and spectral_rolloff algorithms as a tool for Feature Extraction. The algorithms used for making inferences are support vector machine (SVM), K-nearest neighbors, and SVR. The work can be extended even further by combining video data with the audio data for analysis. The system can also be improved by increasing the number of languages it can detect.

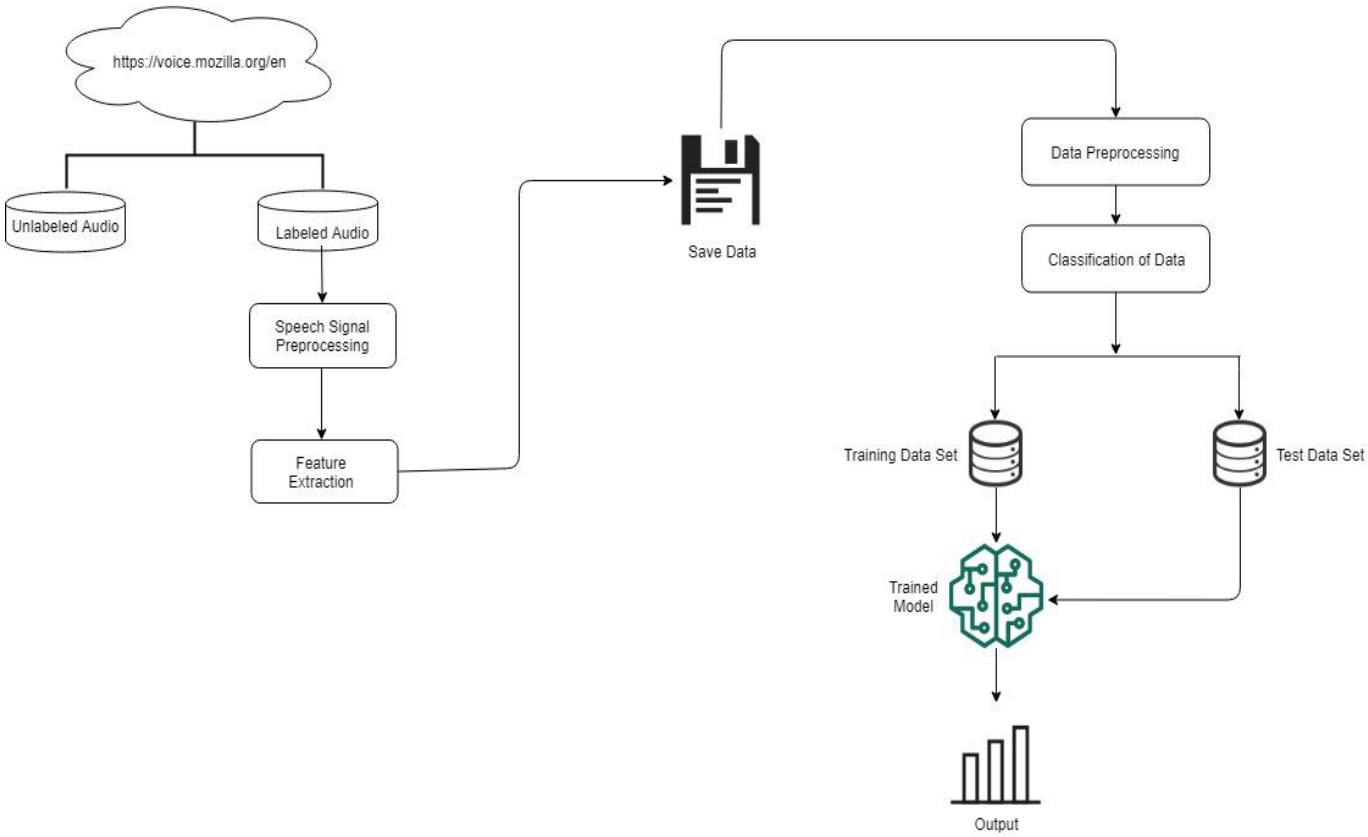
Index Terms - Feature Extraction, Speech processing, age-gender classification, Accents classification, mel-frequency Cepstral Coefficient, zero cross rate, SVM, KNN.

I. INTRODUCTION

Speech identification technology has been applied for various purposes such as voice bio-metrics and assistive technology for disabled people. It has also been used in multiple virtual assistants to identify the speaker. However, speech can differ significantly depending on a number of factors, such as social class, speed, regional dialects, emphasis, and also gender. Variations of these sorts, highly influence the speech recognition system. The vocal features are one of the attributes which are unique to each individual.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Arnav Garg*, Department of Computer Science Engineering, SRM IST, Kanchipuram Tamil Nadu, India.
Kushal Agrawal, Department of Computer Science Engineering, SRM IST, Kanchipuram Tamil Nadu, India.
Mrs. P.Akilandeshwari, Department of Computer Science Engineering, SRM IST, Kanchipuram Tamil Nadu, India.

A wide range of characteristics can be obtained from the human voice. Extracting and Analyzing these attributes can prove to be of great benefit to the security industry.
The main challenge in studying audio data is the extraction process. This process can be cumbersome on IT resources like CPU, memory, and storage as the dataset belongs to the category of big data.
Therefore we are using google cloud platform for this heavy processing. All the attributes obtained would not be of equal importance to the analysis process. It is essential to properly understand each of these attributes and figure out their level of significance.
For analysis purposes, we would be using various machine learning algorithms. A detailed comparison has to be performed to select the best model. In our study, we are going to compare SVM, SVR, KNN, and Naive Bayes. For our models to get excellent precision, we will have first to perform feature scaling the parameters to ensure that each attribute contributes equally to the prediction process.The proposed system can be used in security agencies. It can also be used in police interrogations. This system can also be used by telephone agencies to see which section of the market they are currently leading in and which part of society they need to focus more on.



II. METHODOLOGY

End-point Detection

Whenever we provide an input voice signal to the model, the first step is to locate the starting and ending point of the vocal signal. The main reason for this step is to get rid of the silence which exists before and after a speech. This ensures that the analytical processing focuses on the most valuable section of the sound. In this system, to overcome such issues the algorithm we are going to be implementing is the zero crossing rate, ZCR rule. The information gained from the ZCR algorithm shows the number of times the sound sequence changes its sign per frame. It is given by

$$Z(n)=\frac{1}{2}\sum_{m=1}^{N}\left|\operatorname{sgn}[x(m+1)]-\operatorname{sgn}[x(m)]\right| \quad (1)$$

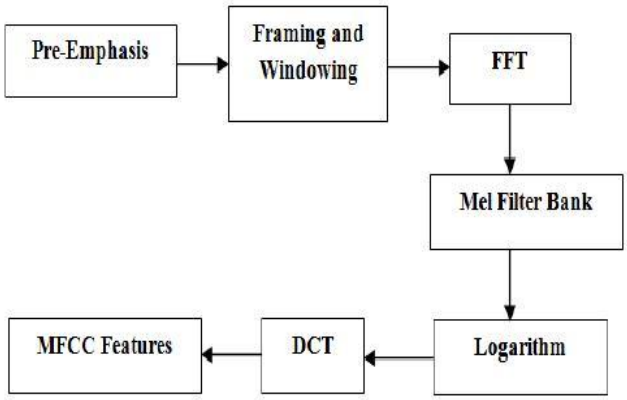
where:

$$\operatorname{sgn}[x(m)]=\begin{cases}+1 & x(m\geq 0)\\ -1 & x(m<0)\end{cases} \quad (2)$$

Mel-frequency Cepstral Coefficient (MFCC)

One of the main feature used is the MFCC. It is a depiction of the short-term power spectrum of a sound. This technique of feature extraction aims to obtain an observation that is highly compact, most suitable, and least redundant for statistical modeling. It is based on the linear cosine transform of a log power spectrum.

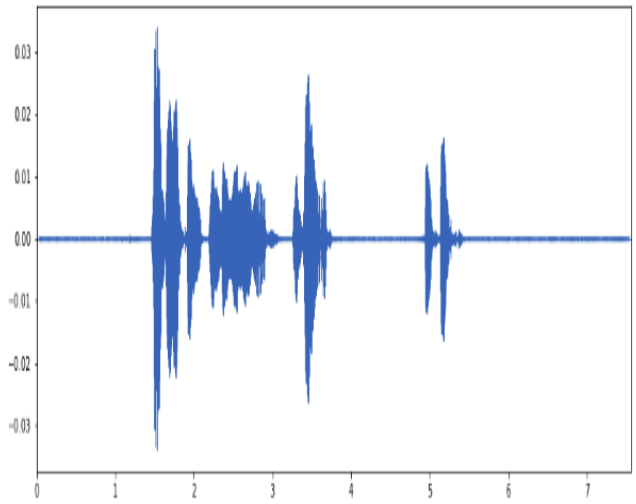
This means that at low frequencies, linear filtration of space takes place and at high frequencies, logarithmic filtration of space takes place. The steps used in the implementation of MFCC are as shown below.



Other features extracted

Chroma stft stands for Short-time Fourier transform and chroma features. Chroma features closely related to the twelve different pitch classes in music. Fourier transform is used to convert time-dependent vocal signals to frequency-dependent vocal signals.

Spectral centroid tells us where the center of mass of the spectrum is placed. Spectral bandwidth is the difference between the upper and lower cutoff frequencies. Spectral roll-off point is the portion of bins in the power-spectrum, at which lower frequencies account for 85% of the power.



Cloud Computing

Scaling up a physical pc can be very expensive and time-consuming. The cloud platform lets us create a customizable machine in a matter of seconds. Cloud computing is on-demand availability of computer system resources. This environment can be scaled as and when required. The cloud mechanism has the right amount of robustness and has measures in case of failure. The charges are minimal in comparison to buying a physical PC. Cloud also offers dynamic scalability. This means that resources automatically increase or decrease depending on the usage. Google cloud platform was used for our purpose. It can be used to create virtual servers with customizable resources. It was used for the heavy processing of the training and testing dataset.

Naive Bayes

It is a straightforward machine learning algorithm that works on a probabilistic model. It is a classification algorithm. It works on the simple principle that every pair of features being classified is independent of each other. Naive Bayes classifier is highly scalable. It has several parameters that can be tweaked.

$P(h|d) = (P(d|h) * P(h)) / P(d)$

Support Vector Machine (SVM)

It is a supervised machine learning algorithm. SVM can perform really well with a linearly and non-linearly seperated dataset. It also produces a good result if the data is limited. The main goal of the SVM algorithm is to search for a hyper-plane in an N dimensional space, where 'N' stands for the number of features used for predictive analysis, that distinctly classifies the data points.

Support Vector Regression (SVR)

A version of SVM which is used for regression problems, was proposed in 1996 by Harris Drucker, Linda Kaufman, Christopher J. C. Burges, Vladimir N. Vapnik, and Alexander J. Smola. This strategy is known Support Vector Regression (SVR). The model created using the support vector arrangement depends only on a small section of the training data, on the grounds that the cost capacity for training the model does not depend on training data that lies on the exterior side of the margin.

K-Nearest Neighbors (KNN)

KNN algorithm is designed to be used for both regression and classification predictive problems. In any classification problem the output is a class member. KNN algorithm is very efficient for our use case as - 1.Calculation time is low 2.Predictive power is good 2.Easy to interpret output.

III. DATA SET

The dataset used for our purpose was taken from an open source Initiative taken by mozilla. The dataset has over four thousand hours of validated voice. The dataset was created by collecting voices of people of different age groups and having different accents. The dataset contains voices from both genders. This dataset mainly contains unstructured data which needs to be studied very closely to obtain a good understanding in order to make predictions.

IV. RESULTS

SVM and KNN algorithms gave the best results in our case. The results for each feature is as below:-

ACCENT (KNN performed better than SVM for this usecase)

96.05911330049261

	precision	recall	f1-score	support
african	1.00	0.96	0.98	24
australia	0.95	0.97	0.96	93
bermuda	1.00	1.00	1.00	3
canada	0.98	0.93	0.95	102
hongkong	1.00	1.00	1.00	4
indian	0.95	0.96	0.95	91
ireland	0.96	1.00	0.98	22
malaysia	1.00	1.00	1.00	3
newzealand	0.94	1.00	0.97	16
philippines	1.00	0.88	0.93	8
scotland	0.94	1.00	0.97	31
singapore	1.00	1.00	1.00	1
southatlandtic	1.00	1.00	1.00	2
wales	1.00	1.00	1.00	6
micro avg	0.96	0.96	0.96	406
macro avg	0.98	0.98	0.98	406
weighted avg	0.96	0.96	0.96	406

AGE (SVM performed better than KNN for this usecase)

92.88025889967638

	precision	recall	f1-score	support
10.0	1.00	0.80	0.89	40
20.0	0.86	0.99	0.92	205
30.0	0.94	0.92	0.93	148
40.0	0.98	0.88	0.93	74
50.0	0.99	0.90	0.95	105
60.0	1.00	1.00	1.00	37
70.0	1.00	0.78	0.88	9
micro avg	0.93	0.93	0.93	618
macro avg	0.97	0.90	0.93	618
weighted avg	0.93	0.93	0.93	618

Gender (KNN performed better than SVM for this usecase)

98.54132901134521

	precision	recall	f1-score	support
female	0.98	0.96	0.97	166
male	0.99	0.99	0.99	448
other	1.00	1.00	1.00	3
accuracy			0.99	617
macro avg	0.99	0.99	0.99	617
weighted avg	0.99	0.99	0.99	617

V. CONCLUSION

MFCC is a one of the most important feature extracted from audio data for statistical modeling. It is one of the most compact and least redundant feature. Though all the algorithms i.e. Naive Bayes, SVM, KNN and SVR are equally useful for analyzing features of audio data, SVM and KNN are more accurate than the others. KNN efficiency can be improved even further by adding a weight to the nearest neighbour.

REFERENCES

1. Speech Signal Feature Extraction Based on Wavelet Transform by Xiaolan Zhao, Zuguo Wu, Jiren Xu, Keren Wang at International Conference on Intelligent Computation and Bio-Medical Instrumentation 2011.
2. MFCC Based Robust Features for English Word Recognition by N.N.Lokhande, P.S.Vikhe , IEEE Conference 2012.
3. Gender- dependent Feature Extraction For Speaker Recognition by Thomas Fang Zheng and Lantian Li, IEEE Conference 2015.
4. Feature Extraction Analysis on Indonesian Speech Recognition System by Adiwijaya, Widi Astuti and Untari N. Wisesty, 3rd International Conference on Information and Communication Technology (ICoICT) 2015.
5. Lpc And Lpcc Method Of Feature Extraction In Speech Recognition System by Divya Gupta and Harshita Gupta, IEEE Conference 2016.
6. Improving Speech Recognition Using Limited Accent Diverse British English Training Data With Deep Neural Network by John H. L. Hansen, Maryam Najafian, Saeid Safavi, Martin Russell, IEEE International Workshop On Machine Learning 2016.
7. Speech Disorder Recognition using MFCC by Prajacta Nagraj, Gunjan Jhawar, and P. Mahalakshmi, International Conference on Communication and Signal Processing 2016.
8. Optimal Feature Extraction and Selection Techniques for Speech Processing by Mukesh A. Zaveri, Ankita N. Chadha, and Jignesh N. Sarvaiya, International Conference on Communication and Signal Processing 2016.
9. Musical Instrument Identification using MFCC by Monica S.Nagwade, Varsha R. Ratnaparkhe, IEEE International Conference On Recent Trends in Electronics 2017.
10. Sound based Human Emotion Recognition using MFCC & Multiple SVM by M.U.Inamdar, Anagha Sonawane, Kishor B. Bhangale, IEEE Conference 2017.

AUTHORS PROFILE

Arnav, is a B-Tech Student at SRM Institute of Science an Technology. His major is computer science. He has conducted multiple research in his selected field. In his free time he contributes in open source projects. His research interests include Machine Learning, Neural networks, Big Data, Cloud computing and Artificial Intelligence.

Kushal, is a B-Tech Student at SRM Institute of Science an Technology. His major is computer science. He has a strong hold on multiple programming languages. His research interests include Cloud computing, machine learning, Neural networks and Digital Security.

Akilandeswari is an Assistant Professor in the School of Computing, Department of Computer Science at SRMIST, received her Master of Engineering in Computer Science from Anna University in 2006, Her research interests include Cloud computing, Artificial Intelligence, machine learning, Neural networks, currently working in uncertainty factors that affects cloud scheduling