INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCH, VOLUME 10, ISSUE 2, FEBRUARY-2019
ISSN 2229-5518

Gender Recognition Using Fast Fourier Transform With Ann

Shantam Vijayputra, Dr. Nalini N

Abstract— This paper deals with an efficient method for gender detection based on the audience's voice in a natural environment. Some features such as Third quartile, entropy, mean frequency act as a key feature; these features are then used to train Artificial Neural Network architecture to classify two different genders (Male and Female). The test result shows that the new method ANN architecture which can analyze and learn better and faster.

Index Terms— Artificial Neural Network, Data Processing, Fast Fourier Transform, Gender Identification, Classification Problem, Supervised Algorithm, Backpropagation.

1. INTRODUCTION

This paper is about a stepwise walkthrough to solve a problem which can help us to detect the gender of the person using a recorded audio file (The audio file must contain the voice of one person at a time). this process involves basics of signal processing which helps in signal transformation and extraction of important wave features and these feature set will be used in the process of training classifier using Artificial Neural Networks and also gives a brief idea about how multi-layer perceptron helps in the creation of a hypothesis that helps in generalizing a classification problem.

This paper comprises of 5 major sections, which will guide you step by step to the research and findings.
These are:

1) Abstract
2) Introduction
3) Research Elaborations
4) Results or Finding
5) Conclusions

2. GENERAL AUDIO CLASSIFIER APPROACH

Audio classification is a supervised problem which has input features and knows output labels, for instance using audio for gender detection consist of only two labels i.e. male voice and female voice but same kind of problem can have more than two output labels for e.g. audio genre classification which contains more than two labels such as pop music, rock music, jazz, Blues, Grunge etc. to solve such problems there are traditional approaches which are given below

1) Data loading and transformation.
2) Feature Extraction from raw data.
3) Data pre-processing.
4) Understand the neural network approach to train and optimize.

2.1 DATA LOADING AND TRANSFORMATION

Now that we have a general idea about the steps involved, it is important that we proceed precisely because even a small error in a single step may lead to compromising our research.

This flow chart explains the very first step i.e. data loading, the audio data comes in a different format which cannot be processed such as wav,mp3 etc. so we need to convert that file to numeric format such that it will be easy to compute.

Now, that the audio is loaded we need to take a step forward to process the audio data, as we know the audio is form of wave and to extract data directly from wave may leads to loss of important data hence required a special type of method to process such as signal processing, which takes input one signal and outputs another transformed signal and for such

Processing we have a very famous algorithm called FFT also known as Fast Fourier Transformation.

Fast Fourier Transformation

A **fast Fourier transform** (**FFT**) is an algorithm that samples a signal over a period of time (or space) and divides it into its frequency components. These components are single sinusoidal oscillations at distinct frequencies each with their own amplitude and phase [1]. This transformation is illustrated in

Figure 1. Over the time period measured in the diagram, the signal contains 3 distinct dominant frequencies.

Equation:

$$\sum_{n=0}^{N-1} a_n e^{-\frac{2\pi i n k}{N}}$$

$$= \sum_{n=0}^{\frac{N}{2}-1} a_{2n} e^{-\frac{2\pi i (2n) k}{N}} + \sum_{n=0}^{\frac{N}{2}-1} a_{(2n+1)} e\text{^}(-2\pi i (2n+1) k/N)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} a_n^{even} e^{-\frac{2\pi i (n) k}{N/2}} + e^{-2\pi i k/N} \sum_{n=0}^{\frac{N}{2}-1} a_n^{odd} e^{-\frac{2\pi i n k}{N/2}}$$

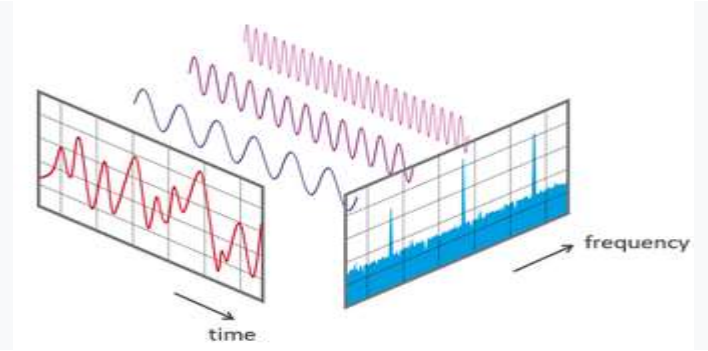

Figure 1: Frequency and time variation

In DSP we convert a signal into its frequency components so that we can have a better analysis of that signal. Fourier Transform (**FT**) is used to convert a signal into its corresponding frequency domain. The below-plotted graph is of male and female voice data before FFT transformation.

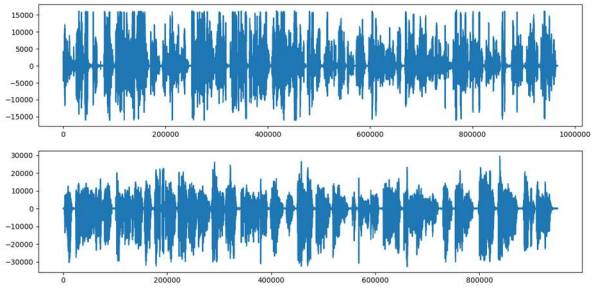

Figure 2: Male vs Female Audio before FFT

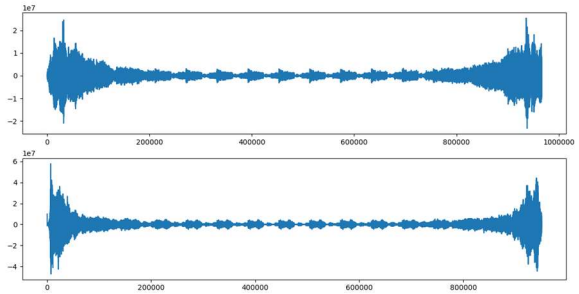

Figure 3: Male vs Female audio after FFT

As we can observe in both the graph we can clearly observe that the transformed signal is much more likely to be predictable the pre-transformed signal.

Now, the transformed signal is ready to act as an input to extract the most common features in the audio set.

2.2. FEATURE EXTRACTION FROM RAW DATA

Here comes the most crucial step for the research i.e. is feature extraction. Feature extraction starts with an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps [2]. The most used feature set in audio classification is:

- MFCC
- LPC
- Spectral Properties

2.2.1 MFCC – Mel Frequency Cepstrum Coefficients [2][9][8] (MFCC) is the most used method that is used in the process of feature extraction.] It is based on the frequency domain which is based on Mel scale based on a human range of hearing. MFCCs, being frequency domain features, are more accurate than time-domain features. MFCC represents the real cepstral

of windowed short time signal which is derived from Fast Fourier Transform (FFT). These coefficients are robust and reliable for variations of speaker and operation environment.

2.2.2 LPC – Linear Predictive Coding (LPC) is a tool most widely used for medium or low bit rate coder. Digital signal is compressed for efficient transmission and storage. Computation of the parametric model based on the least mean squared error theory is known as linear prediction (LP). The signal is expressed as a linear combination of previous samples. Format frequencies are the frequencies where resonance peak occurs.

2.2.3 SPECTRAL PROPERTIES - analysis we can say or **Spectrum** analysis is analysis in terms of a **spectrum** of frequencies or related quantities such as frequency, standard deviations etc. In this paper we will try to use this approach and will use below set of features:

1. Mean Frequency
2. Standard Deviation
3. Median
4. Third Quartile (Q75)
5. First Quartile(Q25)
6. Inter Quartile(IQR)
7. Skewness
8. Kurtosis
9. Spectral Entropy
10. Spectral Flatness
11. Mode

Mean Frequency:
The Mean frequency is a pitch measure that passes the center of distribution of power across the frequencies.

Standard Deviation:
It is defined as the quantity by how many members of the group differs from the mean value of the group.

Median:
It is defined as the value relating to the midpoint of frequency distribution, such that there are equal probabilities of falling up or falling down.

Third Quartile:
It is defined as a number in data for which the 75% of data is less than that number. The third quartile (Q75) is same as the median of the part of the data which is greater than the median. Same as 75 percentile.

First Quartile:

It is defined as a number in data for which the 25% of data is less than that number. The third quartile (Q25) is same as the median of the part of the data which is greater than the median. Same as 25 percentile

Inter Quartile:

It is defined as the difference between the Third Quartile and First Quartile i.e. (Q75-Q25).

Skewness:

Skewness is asymmetry in a statistical distribution, in which curve appears to be distorted or skewed either to the left or right.

Kurtosis:

It is defined as the sharpness of the peak of a frequency-distribution curve.

Spectral Entropy:

Spectral Entropy defines as the complexity of the system. Can be calculated by calculating the spectral power of the signal via squaring its amplitude and normalizing by the number of bins.

Spectral Flatness:

It is defined as the ratio of the Geometric mean to the Arithmetic mean of the magnitude spectrum.

Mode:

It is defined as the data which occurs most of the time in the set of data's considered to be the mode of the data.

3. DATA PREPROCESSING

The data pre-processing is the technique of presenting the data into the required format for the better understanding and the better results but it may differ from situation to situation and also in terms of data to data.

Different data requires different kinds of approach for the pre-processing such as for numerical data the pre-processing techniques can be "scaling", "reduction techniques", whereas in case if the string or character data the pre-processing techniques varies such as "label encoding", "one hot encoding" etc.

In short data pre-processing is the techniques to represent the data or formatting the data based on required circumstances

In this data set, three types of pre-processing techniques have been used such as:

3.1 LABEL ENCODING:

Label Encodings are the alternate approach to encoding categorical values. Label Encoding is simply mapping categorical value to a unique defined integer value. For e.g. "male" or "female" can be mapped to 0 or 1 respectively

.

3.2 NAN VALUES:

Sometimes columns may have missing values in the dataset which may cause an error during execution. To deal with this type of situation we may substitute the mean values of the column or the most occurring elements in the columns. In most of the cases taking average turn out to be the most optimal solution.

Equation:

$$D[i] = \sum_{i=0}^{n} D_i/\mathrm{N}$$

FEATURE SCALING:

Feature scaling is one of the most famous techniques of data pre-processing. Sometimes larger values in data set may take more times to compute to avoid this problem scientist came up with the solution of feature scaling .the most famous one is "MIN MAX SCALER", which is computed by subtracting the minimum value of the column and making ratio of that value to the difference of maximum to the minimum value of the column.

Minimax Scaler:

$$D[i] = (D[i] - Min(D))/((Max(D) - Min(D))$$

Standard Scaler:

$$\boldsymbol{D[i] = (D[i] - Mean(D))/STD(D)}$$

9.64E-02	4.73E-01	8.41E-02	6.01E-02	2.05E-01	2.55E-01	3.68E-01	2.08E-01	6.36E-01	5.65E-01	0.00E+00	9.64E-02
1.26E-01	5.05E-01	1.17E-01	7.76E-02	2.16E-01	2.47E-01	6.44E-01	4.84E-01	6.31E-01	5.92E-01	0.00E+00	1.26E-01
1.79E-01	6.76E-01	1.03E-01	3.43E-02	3.86E-01	4.57E-01	8.85E-01	7.82E-01	4.43E-01	5.48E-01	0.00E+00	1.79E-01
5.28E-01	5.55E-01	5.88E-01	3.90E-01	7.16E-01	4.07E-01	3.15E-02	1.61E-03	9.23E-01	8.56E-01	3.00E-01	5.28E-01
4.52E-01	6.27E-01	4.54E-01	3.18E-01	7.08E-01	4.74E-01	2.77E-02	1.73E-03	9.59E-01	9.26E-01	3.72E-01	4.52E-01
4.41E-01	6.31E-01	4.32E-01	2.74E-01	7.23E-01	5.35E-01	5.18E-02	4.77E-03	9.23E-01	8.70E-01	4.02E-01	4.41E-01
5.26E-01	5.79E-01	5.96E-01	3.75E-01	7.06E-01	4.13E-01	4.02E-02	3.00E-03	9.41E-01	9.00E-01	3.08E-01	5.26E-01
5.72E-01	6.03E-01	5.33E-01	4.46E-01	8.20E-01	4.50E-01	3.63E-02	2.06E-03	9.07E-01	8.47E-01	4.58E-01	5.72E-01
4.86E-01	6.16E-01	5.10E-01	3.56E-01	7.19E-01	4.45E-01	2.77E-02	1.53E-03	9.54E-01	9.11E-01	7.83E-01	4.86E-01
4.48E-01	6.40E-01	4.41E-01	3.05E-01	6.90E-01	4.70E-01	3.03E-02	2.08E-03	9.72E-01	9.52E-01	4.18E-02	4.48E-01
5.56E-01	5.53E-01	6.28E-01	4.10E-01	7.54E-01	4.24E-01	2.42E-02	1.46E-03	9.31E-01	8.64E-01	3.44E-01	5.56E-01

Figure 4: Sample dataset after pre processing

4. UNDERSTAND THE NEURAL NETWORK APPROACH TO TRAIN AND OPTIMIZE

We know that our brain computes very fast and effortlessly irrespective of task such as human face recognition, sound recognition. The key behind is that parallel computation. Thousands or even millions of nerve cells called [3][10] **Neurons** are organized to work simultaneously.

Artificial Neural Networks tries to mimic the architecture of brain neurons to use parallel computation to compute millions of task. Before we go deep to understand ANN we must understand the concept of Perceptron because ANN is nothing but multi-layer perceptron.

Perceptron

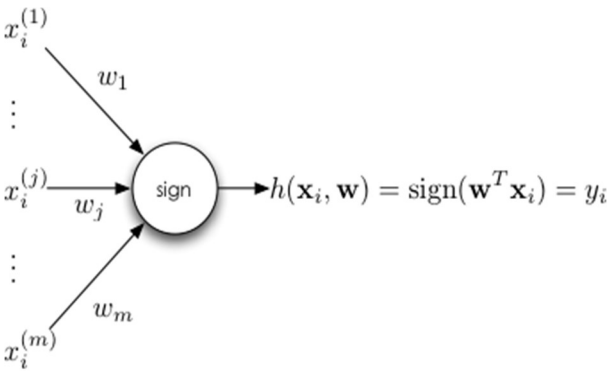

Figure 5: Perceptron Model

Perceptron is called single layer neurons, in neurons, the node is a combination of summation unit and the threshold function. Threshold s and outputs 0 or 1 depending on whether the weighted sum is less than or greater than.

The input node with +ve weights is called excitory and with – ve weights are called inhibitory.

$$D = \sum_{i=0}^{n} w_i x_i$$

If D > T gives output as 1 or else 0.

Multi-Layer Perceptron with Hidden layers

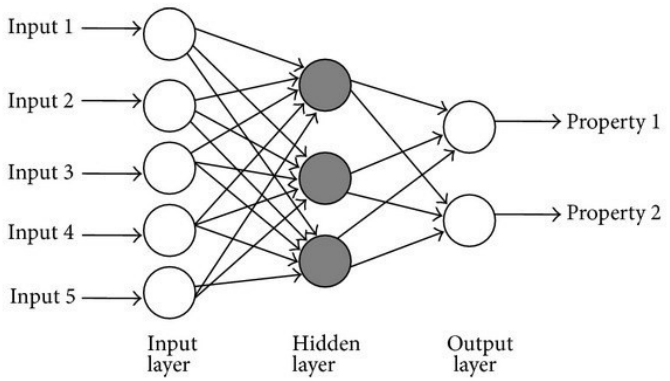

Figure 6: Multi-layer Perceptron

The multi-layer perceptron has one input layer and one output layer the layers in between them is called the hidden layers.

[3] The hidden layers act as a feature detector, they play a critical role in operation for multi-layer perceptron, they gradually discover the salient features that cauterize the training data.
Multi-layer perceptron has two major and most important steps:

- Feed Forward
- Backpropagation

The Backpropagation Algorithm

- Initialize the weights $w_{ij}^{(k)}$ to small random values and choose a positive constant.
- Repeatedly $x_1^{(0)}$…… $x_{M_0}^{(0)}$ to the features of sample 1 to N and cycling back.
- **Feed Forward: for k=0 ….. K-1, compute :**
  $x_j^{(k+1)} = R\left(\sum_{i=0}^{M_k} w_{ij}^{(k+1)} x_i^{(k)}\right)$
  R(s) =$1/(1+e^{-s})$
- **Backpropagation step:** for the nodes in the output layer, j = 1, …. , m_k compute:
  $\delta_j^{(k)} = x_j^{(k)}(1 - x_j^{(k)})\,(x_j^{(k)} - d_j)$
  For layers k = K-1, …… , 1 compute:
  $\delta_j^{(k)} = x_j^{(k)}(1-x_j^{(k)}) \sum_{j=1}^{M_{(k+1)}} \delta_j^{(k+1)} w_{ij}^{(k+1)}$
  For i=1, …… M_k.
- Replace $w_{ij}^{(k)}$ by $w_{ij}^{(k)} - c\delta_j^{(k)} x_i^{(k-1)}$
- Repeat the steps until the weights $w_{ij}^{(k)}$ cease to change significantly.

Using the above algorithm to optimize weights which gives the hypothesis to classification and for our problem it got the learning curve like this:

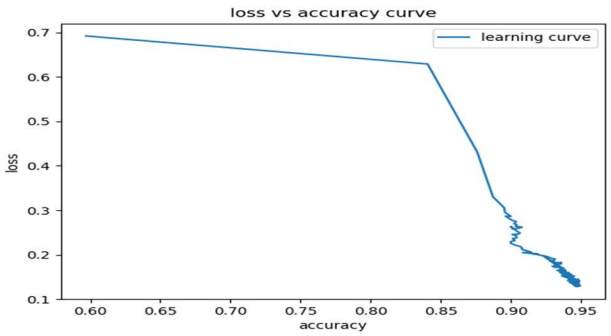

Figure 7: Learning curve

Confusion Matrix:

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known [4].

For the above Model:

cm = [[478, 39] [[40, 489]]

Accuracy:

(Correctly predicted class / total testing class) × 100%

Training: - 94%
Testing: - 92.44%

Precision:

= TP / TP + FP

= 489/ (489+39)
= 0.9261363636363636
= 92.61%

Recall:

= TP/TP+FN

= 489/ (489+40) = 0.9243856332703214 = 92.43%

F1-Score:

= 2*(Recall * Precision) / (Recall + Precision)
=2*(0.9261363636363636*0.9243856332703214)/(0.9261363636363636+0.9243856332703214)
= 0.925252443830496
= 92.52%

5. CONCLUSION

At last, I would like to conclude that ANN is much faster and efficient in terms of learning than other any machine learning algorithms but it requires a quality practice in data set generation and preprocessing because bad quality data may lead to overfitting or underfitting of the model.

6. REFERENCES

[1] Dept. Mathématiques Informatique, Ecole Centrale de Lyon, France {Hadi.harb, liming.chen}@ec-lyon.fr "GENDER IDENTIFICATION USING A GENERAL AUDIO CLASSIFIER"

[2] George Tzanetakis, Student Member, IEEE, and Perry Cook, Member, IEEE "Musical Genre Classification of Audio Signals" 1063-6676/02$17.00 © 2002 IEEE

[3] Book : "Pattern recognition and Image analysis" by Steve jost , Earl Gose, edition: 2000
Isbn: ISBN-81-203-1484-0

[4] Automatic Emotion Recognition from Speech using Artificial Neural Networks with Gender- Dependent Databases ,Firoz Shah.A, Raji Sukumar.A, Babu Anto.P School of Information Science and Technology, Kannur University, Kerala, India firozathoppil@gmail.com, rajivinod.a@gmail.com, bantop@gmail.com

[5] AUDIO CLIP CLASSIFICATION USING LP RESIDUAL AND NEURAL NETWORKS MODELS Anvita Bajpai and B. Yegnanarayana Department of Computer Science and Engineering Indian Institute of Technology Madras, Chennai- 600 036, India {anvita, yegna}@cs.iitm.ernet.in

[6] Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network, The International Arab Journal of Information Technology, Vol. 10, No. 5, September 2013

[7] Gender classification using face image and voice Dharamraj yadav1 Shashwat Shukla2 & Bramah Hazela2 Dept.of Computer Science & Engineering Amity School of Engineering & Technology Amity University, Lucknow campus, India Dept.of Computer Science & Engineering Amity

[8] School of Engineering & Technology Amity University, Lucknow campus, India

[9] Gender Classification Based on FeedForward Backpropagation Neural Network Conference Paper · September 2007 DOI: 10.1007/978-0-387-74161-1_32 · Source: DBLP

[10] International Journal of Computer Applications (0975 – 8887) National Conference on Recent Trends in Information Technology (NCIT 2015) 5 Speech based Gender Identification using Feed Forward Neural Networks

[11] Speech Recognition Using Artificial Neural Network – A Review Bhushan C. Kamble1, Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE) Vol. 3, Issue 1 (2016) ISSN 2349-1469 EISSN 2349-1477 http://dx.doi.org/10.15242/IJCCIE.U0116002 1

7. AUTHORS

- *Shantam vijayputra is currently pursuing bachelors degree program in computer science and engineering in Nitte Meenakshi Institute of Technology, Bangalore,India, E-mail: vshantam@gmail.com*
- *Dr. Nalini N is currently – Professor in Department of Computer Science and Engineering at Nitte Meenakshi Institute of Technology, Bangalore, India, E-mail: nalini.n@nmit.ac.in*