International Journal of Scientific & Engineering Research Volume 9, Issue 3, March-2018
ISSN 2229-5518

SPEECH RECOGNITION USING NEURAL NETWORK FOR CONTROLLING ROBOT

Prof.Lalita Moharkar, Mr.Harsh Sindhwa, Miss.Tejashri Tambe, Miss.Supriya Yenaganti

Abstract— Voices of different people of various ages in a silent and noise free environment by a good quality microphone are recorded. These recorded speech signals are recognized using back propagation algorithm in neural network. Same digits of duration 4 seconds is spoken by these people. These spoken digits are then converted into wave formats. Then features of the recorded samples are extracted by training these signals using LPC. Learning is required whenever we don’t have the complete information about the input or output signal. At the input stage, 210 samples of each digits are applied, then through hidden layers these are passed to output layer. These networks are trained to perform tasks such as pattern recognition. Then this speech signal is used for controlling robot

Index Terms— Linear predictive code(LPC),Neural networks(NN), Back propogation (BP),Self Organizing Map(SOM)

1 Introduction

THE system extracts parameters from the input speech signal to represent vocal characteristics and uses these information to build representative speech model. Speech could be a useful interface to interact with machines. Speech recognition is facing a lot of problems due to variations occurred in speech including the variations because of age ,sex, speed of speech signal, emotions, conditions of speaker can cause the difference in the pronunciation of different persons. In speech recognition process an signal is captured by microphone. These recorded signal was recognized by using back propagation in Neural Network. This recognized speech signal is given to controlling robot.

2 Methodology

2.1 Block Diagram

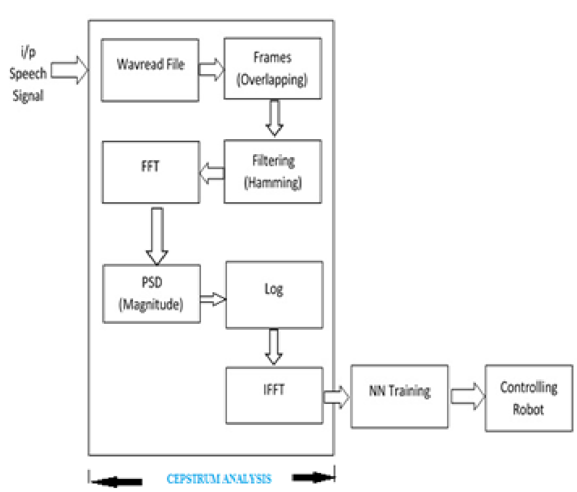

Fig. 1.0 BLOCK DIAGRAM

.

2.1.1 Feature Extraction

Features are an extremely more refined and compressed input to the neural network. So based on theoretical and practical operation, we will extract features from the frames.

2.1.1.1 Capturing the signal

First step in processing part is to capture the signal we need. The signal obtained from human is an analog signal. While storing the speech signal to a computer, the analog signal needs to be digitized first. When people speak something on microphone, this analog signal pressurizes the air and then this analog electric signals goes to the microphone. Analog speech digitization has two steps viz Sampling and Quantification.

2.1.1.2 Silence Removal

The silence signal does not contain any information and is of no use to the speech sysytem. If we do not remove the silence signal it will make the processing of the system larger and consume more time and space while getting information or features from the signal. So only the signal part which contains the actual speech content is useful for recognition

.

2.1.1.3 Normalization

Normalization is a method which adjusts the volume of audio files to a standard level. The normalization of the signal is done in a simple and unique way for example scaling, and offsetting the signal so that it falls between levels -1 and +1.

2.1.1.4 Pre-emphasis

Pre-emphasis is a very simple signal processing method which is used to increase the amplitude of high frequency bands and decrease the amplitudes of lower bands, this process boosts the energy in high frequency which will give more information for the further process and also will improve the signal for same frequency. Then this signal is framed .The width of the signal is generally about 30 mS with an overlap about 20 mS that's mean it's shifted generally 10 mS. Each frame contains N number of sample points of the speech signal. If the frame is short, then it will not get the enough samples to estimate the reliable spectral that is the reason why we need to select the frame signal into 20-40mS frame which will gives throughout.

2.1.1.5 Windowing

An audio signal is an unstable signal, meaning. In a very out a short segment of the speech signal .The time for which the signal is considered for processing is called a window, and the data acquired in a window is called a frame.

2.1.1.6 Linear Predictive Coding

LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. LPC is frequently used for transmitting spectral envelope information, and as such it has to be tolerant of transmission errors. Transmission of the filter coefficients directly is undesirable, since they are very sensitive to errors.

3 Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way of biological nervous system, such as the brain process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing element working in unison to solve specific problems. An ANN is configured for a specific application such as pattern recognition or data classification through a learning process. Neural networks are interesting because of their potential use in prediction and classification problems. Neural networks are nonlinear data driven self adaptive powerful tools for modeling, especially when the underlying data relationship is unknown. In artificial neural networks have been applied successfully to speech recognition, image analysis and adaptive control, in order to construct software agents (in computer and video games) or autonomous robots. Most of the currently employed artificial neural networks for artificial intelligence are based on statistical estimations, classification optimization and control theory.

The cognitive modeling field involves the physical or mathematical modeling of the behaviour of neural systems, ranging from the individual neural level (e.g. modeling the spike response curves of neurons to a stimulus) through the neural cluster level (e.g. modeling the release and effects of dopamine in the basal ganglia) to the complete organism (e.g. behavioural modeling of the organism's response to stimuli). Artificial intelligence cognitive modeling, and neural networks are information processing paradigms inspired by the way biological neural systems process data.

Neural networks have been used in classification problems, speech recognition, stock market prediction and performance analysis, pattern matching, natural language procession, vision processing. Signal processing, image processing, decision making, optimization and control systems.

Characteristics of neural networks:

- **Learn by Example**
- **Mapping Capabilities**
- **Robust Systems**
- **Fault Tolerant**
- **Parallelism**
- **Flexibility**
- **Adaptive Learning**
- **Ability**
- **Self Organization**

3.1 Back Propogation

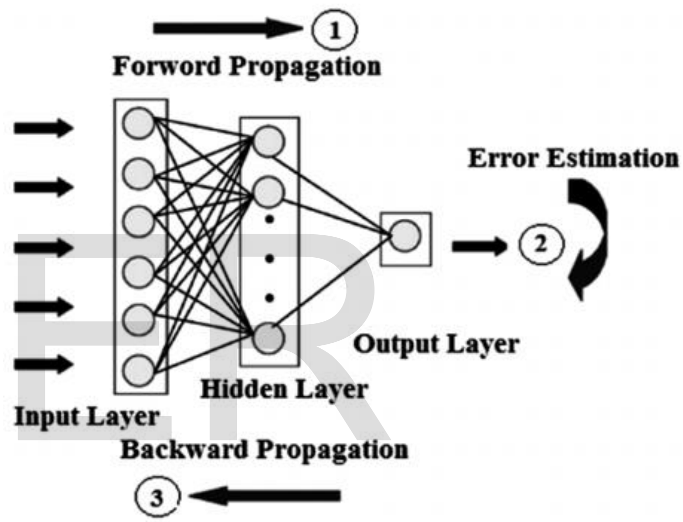

Fig.1.1 Back Propogation

Back Propagation algorithm is used to train feed forward neural. It has two phases that is forward and backward pass phase. In forward pass phase, computes the function signal to get the output for input to hidden layer and hidden to output layer. In backward pass phase, computes the error signal to get the output for backward form.

Algorithm:

Step 1: Initialization

Step 2: Consider at random weight and input value.

Step 3: Do forward pass through the net to produce output.

- I/O applied
- Multiplied with weight
- Summarized (Produced net value)
- 'Squashed' by sigmoid activation function.
- Output passes to each neurons of the next layer.

Step 4: Computes the activation signal for input to hidden and hidden to output layer using step 2.

Step 5: Computes the error over the output neuron by comparing desired with actual output.

Step 6: Use the error to computes the change in hidden to output and input to hidden weights value
Step 7: Update the weight value.
Step 8: Repeat step 3-6 until network is not train.

4 Results

Step 1: We recorded the speech signal of an individual 3 times using normalisedlpc.m file.

Step 2: We took about 150-210 samples of speech signal.

Step 3: We saved the database using Matlab file.

Step 4: Then we trained it using BP algorithm, we took about 150 iterations.

Step 5: We then successfully predicted the speech signal.

```
record voice

P =

     9

record voice

P =

     9

record voice

P =

     9

record voice

P =

     3

record voice

P =

     9
```

Fig.1.2 Extracted digit

```
Iteration   124 | Cost: 1.492074e+000
Iteration   125 | Cost: 1.492030e+000
Iteration   126 | Cost: 1.491648e+000
Iteration   127 | Cost: 1.491543e+000
Iteration   128 | Cost: 1.491415e+000
Iteration   129 | Cost: 1.491334e+000
Iteration   130 | Cost: 1.491297e+000
Iteration   131 | Cost: 1.491187e+000
Iteration   132 | Cost: 1.491148e+000
Iteration   133 | Cost: 1.491107e+000
Iteration   134 | Cost: 1.491063e+000
Iteration   135 | Cost: 1.491029e+000
Iteration   136 | Cost: 1.490983e+000
Iteration   137 | Cost: 1.490932e+000
Iteration   138 | Cost: 1.490916e+000
Iteration   139 | Cost: 1.490896e+000
Iteration   140 | Cost: 1.490856e+000
Iteration   141 | Cost: 1.490814e+000
Iteration   142 | Cost: 1.490677e+000
Iteration   143 | Cost: 1.490416e+000
Iteration   144 | Cost: 1.490095e+000
Iteration   145 | Cost: 1.489991e+000
```

Fig.1.3 Neural Network training output

```
record voice
enter the number 1
record voice
enter the number 2
record voice
enter the number 3
record voice
enter the number 4
record voice
enter the number 5
record voice
enter the number 6
record voice
enter the number 7
record voice
enter the number 8
record voice
enter the number 9
```

Fig.1.4 Record voice as digit

Digit	Accuracy
0	90%
1	80%
2	75%
3	70%
4	70%
5	80%
6	50%
7	50%
8	65%
9	50%

Table 1.0 Accuracy Table

5 Future Scope

The speech recognition system can be easily given to the hardware system. There are various applications which can be controlled by the speech system. Recognized speech can be given to control devices viz lights,fans etc.We will design a wheelchair which will be able to take speech input accurately.

6 Conclusion

In this paper we conclude that neural networks can be very powerful model for the classification of speech signals. Some types of simplified models can recognize the small set of words. The performance of the neural networks is being impacted largely by the pre-processing technique. Although none of the approaches proved to be good enough for practical purposes with the present extent of development, they were good enough to prove that translating speech into trajectories in a feature space works for recognition purposes. The human speech is an inherently dynamical process that can be properly described as a trajectory in a certain feature space. Even more, the dimensionality reduction scheme proved to reduce the dimensionality while preserving some of the original topology of the trajectories, i.e. it preserved enough information to allow a good recognition accuracy. It is interesting to note that despite the fact that the SOM (Self Organizing Map) has been used in the speech recognition field for more than a decade, nobody has used it to produce trajectories, but only to generate sequences of labels. Finally, the new approach developed for training the neural network's architecture proved to be simple and very efficient. It reduced considerably the amount of calculations needed finding the correct set of parameters.If the traditional approach had been used instead, the amount of calculations would have been higher.

7 Acknowlegment

We are grateful to Xavier Institute of Engineering for giving us the opportunity to actually work on a project and give a lot of knowledge. With deep sincere of gratitude the express are sincere thanks to our esteemed and worthy guide Prof. Lalita Moharkar for her valueable guidance in carrying out this work under their effective supervision, encouragement and cooperation. Also thanks to our HOD, Dr. Suprava Patnaik who is always ready to help in our project and our Principle, Dr. Y. D. Venkatesh who is gave us an opportunity to work on project in the collage. Also the express our heartiest thanks to all the Lab Assistance of Xavier Institute of engineering, for their patience and support in the labs.

8 References

[1] ' Dr. Caisimis Kulikowski, Dr. James"Robust Speech Recognition using Neural Network and hidden markov model", October 1999.

[2] 'Xiaoguo Xue,"Joind Speech and Spesker Recognition using Neural Network", May 2013.

[3]'Pankaj Rani, Sushil Kakkar,Shweta Rani"Speech Recognition using Neural Network",International journal of computer applications(0975-8887), 2015.

[4] John Paul Hosom, Ram Ad Mark Fanty, "Speech Recognition using Neural Networks" volume.1, July 6 1999.

[5] Antanas Lipeika, Joana Lipeika, Loimutis Telksnys, "development of Isolated Word Speech Recognition System" volume.30, No.1, PP 37-46, 2002.

[6] Ben Gold and Nelsom Margan, "Speech and Audio Processing" Willey addition New Delhi, 2007.

[7] Wouter Geuaert, Georgi Tsenav, Valeri Mladenov, "Neural Network used for Speech Recognition" Journals Automatic Control, volume.20.1.7, 2010

[8] Santosh K. Gaikwad, Bharti W. Gawali, Pravin Yannawar, "A Review of Speech Recognition Technique" IJCA, volume.10, No.3, November 2010.

[9] Lakshami Kanaka, Venkateswarlu Revada, Yasautcha Kumari Rambatla and Koti Verra NagayaAnde, volume.8, issue. 2, March 2011, IJCST.