RBF SPEAKER RECOGNITION USING MFCC DELTA COEFFICIENTS

P.P.S.Subhashini
M.Tech, R.V.R. &J.C College of Engg., Affiliated to Acharya Nagarjuna University

ABSTRACT: Training neural network is in general, a challenging nonlinear optimization problem. Radial basis function neural networks provides an attractive possibilities for solving signal processing & pattern classification problems. Several algorithms have been proposed for choosing the RBF neural network prototypes & training the network. The selection of the RBF prototypes & the network weights are the system identification problem. Various derivative based & derivative free methods have been used to train the neural networks. The proposed thesis implements an enhanced feature extraction method for speech signal to train the RBF neural network based. The implementation is then compared with the early existing Back Propagation neural networks for the analysis. The proposed work is tested on the speaker Recognition problem on TIMIT dataset. It is shown that the use of RBF Neural Network and MFCC Delta coefficients results in better classification & faster learning than Back Propagation neural networks.[1]

I. INTRODUCTION

Speaker recognition is the identification of the person who is speaking by characteristics of their voices (voice biometrics), also called voice recognition. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on specific person's voices or it can be used to authenticate or verify the identity of a speaker as part of a security process. Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. These acoustic patterns reflect both anatomy (i.e., size and shape of the throat and mouth) and learned behavioral patterns (i.e., voice pitch, speaking style.) Speaker verification has earned speaker recognition its classification as a "behavioral biometric"1, 2. A radial basis function neural network is trained to perform a mapping from an m-dimensional input space to an n-dimensional output space. Each speaker recognition system has two phases: Enrollment and verification. During enrollment, the speaker's voice is recorded and typically a number of features are extracted to form a voice print, template, or model. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print. For identification systems, the utterance is compared against multiple voice prints in order to determine the best match (es) while verification systems compare an utterance against a single voice print. Because of the process involved, verification is faster than identification.This paper focuses on speaker recognition using Radial Basis Function Neural Network based on MFCC Delta coefficients . From the results it was observed that the proposed method has very high success rate in recognizing different speaker identities. Mel-frequency warping Human ear perception of frequency contents of sounds for speech signal does not follow a linear scale. Therefore, for each tone with an actual frequency f, measured in Hz, a subjective pitch is measured on a scale called the „mel" scale. The mel frequency scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000Hz. To compute the mels for a given frequency f in Hz, a the following approximate formula is used.

Mel (f) = Sk = 2595*log10 (1 + f/700)

The subjective spectrum is simulated with the use of a filter bank, one filter for each desired mel-frequency component. The filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval.

Cepstrum

In this final step, we convert the log mel spectrum back to time. The result is called the Mel Frequency Cepstrum Coefficients (MFCC). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients (and so their logarithm) are real numbers, we can convert them to the time domain using the discrete cosine transform (DCT). By doing DCT, the contribution of the pitch is removed. In this final step Log Mel spectrum is converted back to time. The result is called the Mel Frequency Cepstrum Coefficients (MFCC). The discrete cosine transform is done for transforming the mel coefficients back to time domain. [10]

$$\tilde{C_n} = \sum_{k=1}^{K} (\log \tilde{S}_k) \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \cdots$$

where n=1,2,....K

Whereas S_k, K = 1, 2, … K are the outputs of last step.

A simple way to compute deltas would be just to compute the difference between frames. Thus the delta value d(n) for a particular cepstral value c(n) at time t can be estimated

Design Approach

This project aims towards the implementation of an multi input multi output Radial basis function (RBF) network for the enhancement of the computational effort required for training the network.The test are carried out for speaker recognition problem. For each speech signal 13 Mel Frequency Cepstral Coefficients (MFCC) technique & is classified into one of 3 speakers.The RBF network is trained with four utterances for each word and tested with three more utterances which are different from the trained utterances.[11]. The reformulated RBF networks were trained using the hidden layer function of equation g $(\nu)=[g_0(\nu)]^{1/(1-p)}$ and with the linear generator function of

$g_0(v)=av+b$,with a=1 and b=1. The exponential parameter p was taken as 2 in the simulation results presented here. The training algorithms were initialized with prototype vectors randomly selected from the input data, and with the weight matrix W set to 0.After some experimentation, it was concluded that gradient descent optimization algorithm was terminated when the error function of equation $E=1/2\|Y-Y^\wedge\|^2$ decreased by less than 0:1%. The performance of training method was explored by averaging its performance over five trials, where each trial consisted of a random selection of training and test data.The number of hidden units in the RBF network was varied between 1 and 15.

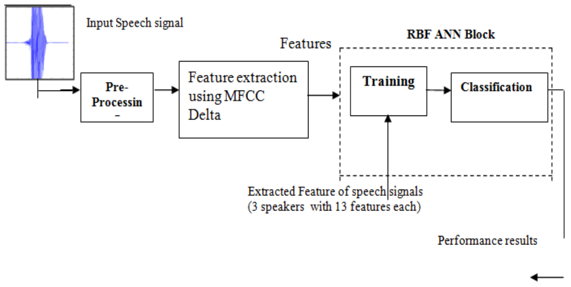
Input Speech signal
Features
RBF ANN Block
Pre-Processin
–
Feature extraction using MFCC Delta
Training
Classification
Extracted Feature of speech signals
(3 speakers with 13 features each)
Performance results

Fig. 1 Speaker Recognition Blok diagram

Results and Analysis

The plot of one of input speech signal considered is shown in the Fig 2.

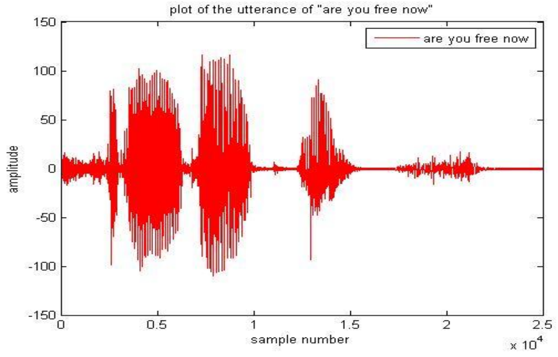

Fig. 2 Speaker Recognition Blok diagram

The Graph in the Fig.2 Shows Percentage of correct classification for the applied test samples by varying number of hidden units. It is observed that this method is successful in recognizing the speaker with correct classification rate of 88 %.

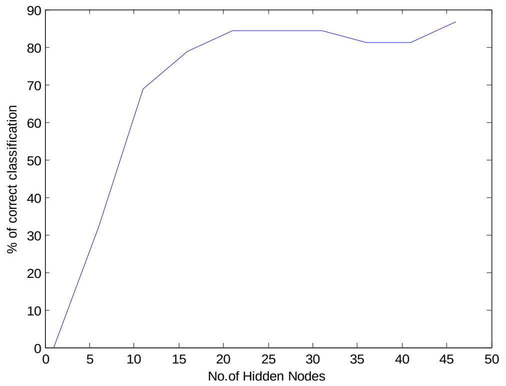

Fig. 3 Percentage of Correct Classification Versus The no. of hidden nodes

It can be seen from the above Fig.3 the Mean square Error versus number of Iterations for training the RBF neural network. It was observed that Mean Square Error gradually reduces from the value 1, as numbers of iterations are increased to and is terminated when the error function decreases by less than 0.0166.

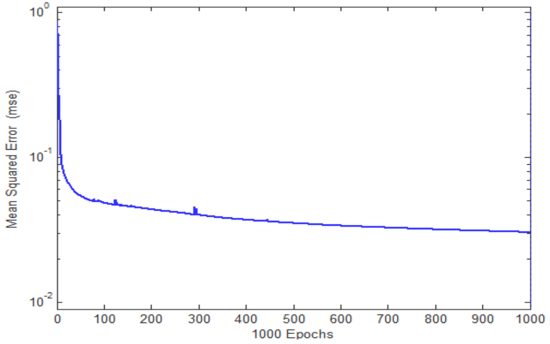

Fig. 4 Mean square Error Versus No. of Epochs

II. CONCLUSIONS

The experiments reported in this paper verify that the RBF speaker Recognition using MFCC Delta feature extraction has reduced computational effort. But Back propagation training requires more iterations for convergence

Future Scope:

Further research could focus on the application of Kalman filter training to RBF networks with alternative forms of the generator function. In addition, the convergence of the Kalman filter could be further improved by more intelligently initializing the training process. Other work could focus on applying these techniques to large problems to obtain experimental verification of the computational savings of decoupled Kalman filter training.Better results can be obtained with the advanced kalman filters like IKF.

Keywords

RBF –Radial Basis Function

BP - Back Propagation

MFCC- Mel Frequency Cepstral Coefficients

Tables: The experimental values for number of iterations are tabulated below

Table 1: No of iterations for hidden nodes from 1 to 15

hidden Nodes training methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Back	79	89	44	68	33	87	69	99	65	47	61		50	39	38
Prppagation												65			
RBF	8.5	9.5	7	6	9	8	8	9	11	10	9	9	9	10	8

Table2**:** The Values of Percentage of correct classification with hidden nodes varying from 1 to 15 are shown in table 2.

hidden nodes Training Methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Back Prppagation	72	71	70	71	72	70	72	70	72	68	73	75	68	73	75
RBF	69	71	64	57	46	54	42	45	68	48	48	54	42	45	68

REFERENCES

[1] Haykin, “Neural Networks A comprehensive Foundation” ,Pearson Education

[2] Jacek M.Zurada, “Introduction To Artificial Neural Networks” , A Jaico Book

[3] B. Yagnanarayana, “Artificial Neural Networks”, Prentice-Hall-India.

[4] Hassoun, ”Fundamentals of Artificial Neural Networks”,A Jaico Book

[5] Satish Kumar, ”Neural Networks”,Tata –McGraw –Hill

[6] Rabiner LR, juang BH, “Fundamentals of speech recognition”, Prentice Hall India

[7] Sadaoki Furui ,Furui Furui, “Digital Speech Processing, Synthesis, and Recognition”

[8] M R Schroeder, ”Speech and speaker Recognition “

[9] John L Ostrander, Timothy D ,”Speech Recognition Using LPC Analysis”

[10] Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk “Speech Recognition using MFCC” International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand)

[11] MFCC and its applications in speaker recognition” Vibha Tiwari, Deptt. of ElectronicsEngg., Gyan Ganga Institute of Technology andManagement, Bhopal, (MP) INDIA (Received 5 Nov.,2009, Accepted 10 Feb., 2010).