Recognition of Voice signals for Oriya Language using wavelet Neural Network

S. Mohanty
Dept of Computer Science
Utkal University
Bhubaneswar

S.Bhattacharya
C.V.Raman College of Engg.
MIE, Bhubaneswar-10
suman1_bh@yahoo.com

ABSTRACT

Speech recognition is both speech oriented and speaker oriented. Both have fuzzy effect on them which adds to the hardness. During speech recognition, separation of the words and again redundancy of the voice responsible for the creation of words due to the vowels makes it difficult for analysis. We are trying with the wavelet neural network model to make an effective analysis for the recognition of speech signals. Here the main characteristics of speech/voice like frequency, intensity, accent and quality are analyzed for better output limiting the associated noise. Here we have used the concept of "wavelon" and "scalon". Using the control parameter of the network we have designed the wavelet network for two characters at the beginning and also we are extending for the rest of the character.

1. INTRODUCTION

Human speech presents a formidable pattern classification task to recognize systems. The speech signal is extraordinarily complex. The human recognizes the speech by recognizing several types of cues-the predominant cues are acoustic. Chief among the acoustic cues are the frequency content of the speech waveform , and the time dependent changes in that frequency content. In general , in its most simplistic form speech can be viewed as a stochastic process involving two principal dimensions - time and frequency. The complexity of the speech recognition task lies in the fact that a given utterance can be represented by an effectively infinite number of time- frequency pattern. A human speech signal is produced by moving the vocal tract articulators towards target positions that characterize a particular sound. Since these articulatory motions are subject to physical constraints that vary from subject to subject and since they are stochastic in nature they do not produce consistently clean identifiable phonetic targets in the train of speech. Instead these articulations from acoustic- phonetic trajectories that have a high degree of variability in both time and frequency domains. Connectionist speech processing and recognition systems are well suited.

2. ARCHITECTURE

The wavelet transform theory and its implementation are limited to the wavelets of the small dimension, where as the ANN are the powerful tools for handling the problems of higher dimensions. Combination of both results in Wavelet network.The weakness of each other compensates and it can handle problems of larger dimension and it also shows efficient network construction methods.

The wavelet network can approximate function ψ:R $\rightarrow$ R There exitst a denumerable family of the form.

$$\phi=\{dk^{1/2}\psi(dkx-tk);\ tk \in R, dk \in R^{+}\ \&\ k\in Z\}\text{-----------------------}(1)$$

Here dk's and tk's are dilation and translations respectively.

The function can be represented as

$$f'(x)=\sum_{k=1}^{N} w_k\ d_k^{1/2}\ \Psi[d_k\ x-t_k\ ]\text{--}(2)$$

Here w_k is the weight associated with each of the wavelons. f'(x) is the approximation of the function f(x) & the accuracy level of the approxinmation goes on in creasing with larger value of N. The function f(x) is otherwise known as wavelet decomposition function.

2.1 Counter Propagation Network (CPN)

In a multi layer feed forward network the training process is slow and its ability to generalize a pattern mapping task depends on the learning rate and the number of units in the hidden layer. On the other hand, a different pattern mapping strategy, namely counter propagation, based on unsupervised learning uses the winner-take-all instar learning for the weights from the units in the input layer to the units in the hidden layer. The CPN provides a practical approach for the pattern mapping task since learning is fast in this network. It consists of four major building blocks - Input Layer, Instar, Competitive network and Out star. The network in the fig consists of two feed forward networks with a common hidden layer. The feed forward network formed by the layers 1,3,2 is used for forward mapping and the network formed by layers 5,3,4 is used for inverse mapping. (if it exists) between the given input-output pattern pairs. Each feed forward network uses a combination of instar and outstar topologies. The first and the second (hidden) layers of a feed forward network form a competitive learning system and the second (hidden) and third layers form an outstar structure. Learning takes place in the instar structure of the competitive learning system to code the input patterns a_I and in the outsrtar structure to represent the output patterns b_l . CPN is generally used for image classification, where the complexities are more.

2.2. Winner-Take- all learning (fig 1)

1. Select an input vector a_l from the given training set (a_l, b_l), l=1,2,…..L.
2. Normalise the input vector and apply it to the CPN competitive layer.

3. Determine the unit that wins the competition by determining the unit k whose vector w_k is closest to the given input.
4. Update the winning unit's weight vectors
   $w_k(m+1)=w_k(m)+\eta(a_l-w_k(m))$
5. Repeat step 1 through 4 until all the input vectors are grouped properly by applying the training vector several times.
   After successfully training the weight vector leading to each hidden unit represents the average of the input vectors corresponding to the group represented by unit.

2.3 Training Outstars of CPN

1. After training the instars apply a normalized input vector a_l to the input layer and corresponding desired output b_l to the output layer.
2. Determine the winning unit k in the competitive layer
3. Update the weights on the connections from the winning competitive unit to the output units as $v_k(m+1)=v_k(m)+\eta(b_l-w_k(m))$
4. Repeat steps 1 through 3 until all the vector pairs in the training data are mapped satisfactorily.

After successful training the outstar weight vector for each unit in the hidden competitive layer represents the average of the subset of the output vectors corresponding to the input vectors belonging to that unit.

3.RESULTS

3.1

In this paper , first a broader view of CPN (Counter propagation Network) is given. Secondly the signal from two different speakers are taken and it is made noise free by the application of it.It extracts a noise from the corrupted input speech signal & gives the desired output signal.In BPN the accuracy level is more but simultaneously the comutational complexities is also more where as in CPN the computational complexities is less but the accuracy is not high .It we used wavelet network, the efficiency increases for processing the large input data & can handle both linear & non-linear system efficiently.

3.2.

The Oriya alphabet (KAA) is taken from two different speakers. It consists of the consonant (KA) and the vowel (AA). During the collection of the signal we also got noise. The noise has to be cancelled before training phase. We have adopted the BPN technique for noise cancellation. After that the error is minimized. Counter Propagation Network and the wavelet are used for further training the signal.

3.3.

The input signal with the noise part is taken in to training phase. (fig 2a). The desired signal is also estimated and shown in (fig 2b).The estimated output is shown in (fig2c) and the error curve is plotted by taking the neuro-wavelet technique (fig 2d). It shows good response . We have also tested it by other algorithms. But the error correction curve for our technique shows a better response. We have plotted the graph for only one speaker. Mainly we have tested the working of the Neuro- wavelet technique and it proves to be efficient for further work. This seems that it will also show good response for many speakers uttering the same syllable.

4. ANN

Artificial neural network takes their names from the network of nerve cells in the brain. It provides an unique computing architecture. Recently ANN has been found to be an important technique for classification and optimization problem. ANN is capable of performing non- linear mapping between the input and output space due to its large parallel interconnection between different layers and the non- linear processing characteristics. An artificial neuron basically consists of a computing element that performs the weighted sum of the input signal and the connecting weight. The sum is added with the bias or threshold and the resultant signal is then passed through a non-linear element of tanh(.) type. Each neuron is associated with three parameters that can be adjusted during learning; these are connecting weights, the bias and the slope of the non linear function.

5. LEARNING OF ANN

The learning of the ANN may be supervised in the presence of the desired signal or it may be unsupervised when the desired signal is not accessible. Rumelhart developed the back propagation algorithm, which is central to much work on supervised learning in multi layer NN. A feed forward structure with input, output, hidden layers and non linear sigmoid functions are used in this type of network. In recent years many different types of learning algorithm using the incremental back propagation algorithm, evolutionary learning using the nearest neighbor MLP and a fast learning algorithm based on the layer by layer optimization procedure are suggested. In case of unsupervised learning the input vectors are classified into different clusters such that elements of a cluster are similar to each other in the same sense. This method is called competitive learning, because during the learning process a set of hidden units compete with each other to become active and perform the weight change. The winning unit increases its weights on those links with high input values and decreases them on those with low input values. This process allows the winning unit to be selective to some input values.

5. CONCLUSION

From the part of the simulation part we have done it is concluded that the Wavelet-Neural network technique works out to be better technique. It results towards an accurate output signal by eliminating noise to an acceptable level from the desired signal. It takes little bit longer time than other processes, but supports non-linear mapping between the input-output pairs. Speech and speaker identification has got wider applications in many corporate office, banks, police etc where secrecy is prime concerned. Better results can be obtained by increasing the sampling frequency. Which involves computational time because more iterations are to be carried out for the same duration of input signal. This can be overcome by the use of high end processors.

References

[1] Rabiner L. and Juang B.H, Fundamental of speech recognition, Prentrece Hall, 1993.

[2] Pati Y C and Krishnaprasad P S, "Analysis and Synthesis of Feedforward Neural Networks using Discrete Affine Wavelet Transform", IEEE Tr. On Neural Networks, vol4, no 2, pp. 73-85, January 1993.

[3] Harold H Szu and Shubha Kadambe,"Neural Network Adaptive Wavelets for Signal Representation and Classification",Optical Engineering , vol.31.no.9, pp.1907-1916, September 1992.

[4] Shashidhara H.L. , Lohani Sumit, and Gadre M Gadre, " Wavelet Neural Network for function Learning", Proceedings of 5th National Conference on Communications (NCC-99), PP.-222-227, Jan 1999.

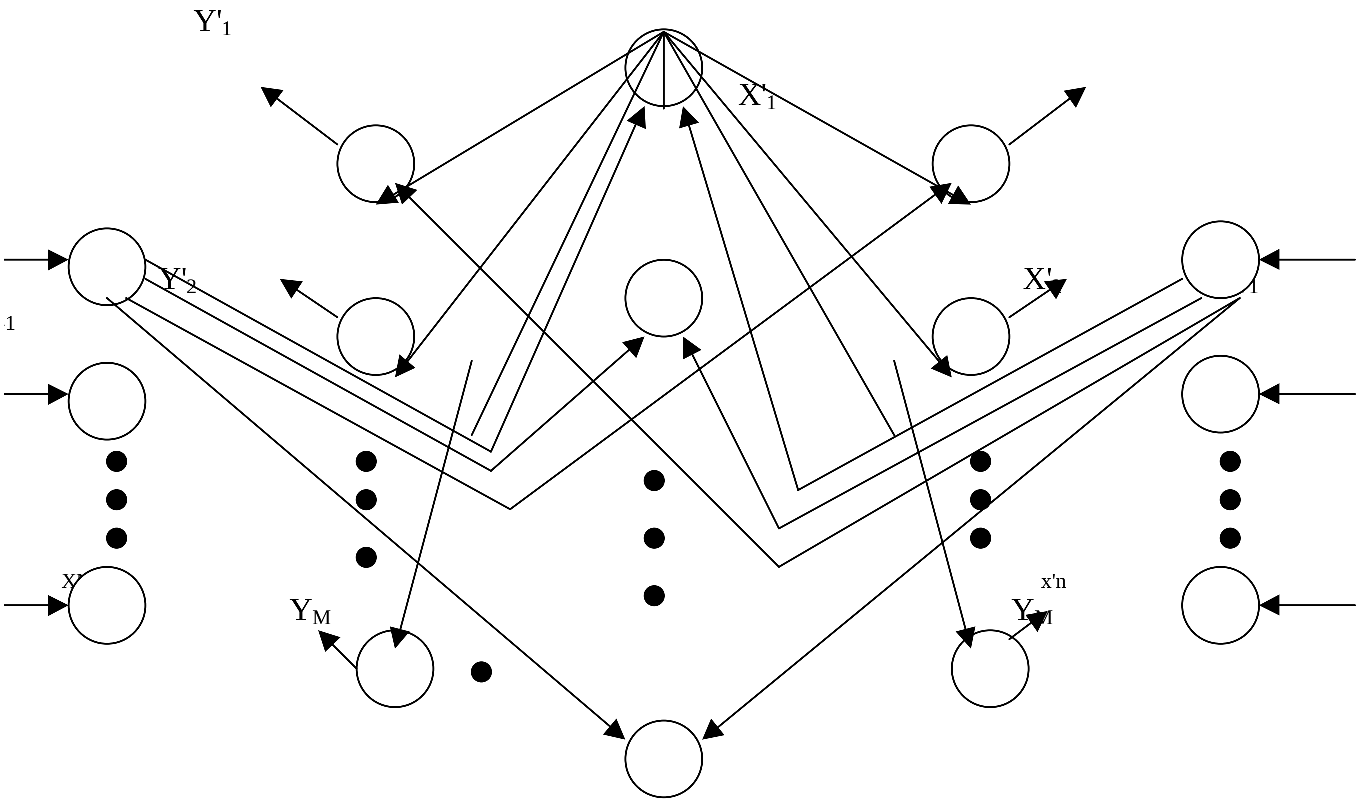

Figure I