Speech Recognition using the Epochwise Back Propagation through time Algorithm

Neelima Rajput
Department of C.S.E
GBPEC, Pauri Gharwal
Uttrakhand, India

S.K.Verma
Department of C.S.E
GBPEC, Pauri Gharwal
Uttrakhand, India

ABSTRACT
In this paper, the artificial neural networks are implemented to accomplish the English alphabet speech recognition. The design an accurate and effective speech recognition system is a challenging task in the area of speech recognition. We implemented a new data classification method, where we use neural networks, which are trained and performance can be defined on the basis of recognition rate. This method gave comparable result to the already implemented neural networks. In this paper, Back propagation neural network architecture used to recognize the time varying input data, and provides better accurate results for the English Alphabet speech recognition. The Epochwise Back Propagation through time (BPTT) algorithm uses the epoch values of input signal to train the network structures and yields the satisfactory results.

Keywords Artificial
Neural Network, Back Propagation Neural Network, Epoch, Speech Recognition.

1. INTRODUCTION
Speech recognition system enables the machine to understand the human speech and react accordingly. It allows the machine to automatically understand the human spoken utterances with the speech signal processing and pattern recognition. In this approach is the machine converts the voice signal into the suitable text or command through the process of identification and understanding. Speech recognition is emerges as a vast technology in current time. It also plays an important role in information theory, acoustics, phonetics, linguistics, and pattern recognition theory and neurobiology disciplines. speech recognition technology become a key technology in the computer information processing technology as there is rapid advancement in the software, hardware and information technology. The features of input audio signal are compared with the voice template stored in the computer database in speech recognition system by using the computer systems. Recognition results are mainly depends upon the matching techniques used for matching the audio signal characteristics. To improve the recognition rate and better recognition results neural networks are used.
A neural network is a powerful tool which used to adapt and represent the complicated input outputs. Neural nets are basically interconnected networks of relatively simple processing units, or nodes that work simultaneously. They are designed to mimic the function of human neuron biological networks. The processing units of neural networks are termed as neurons. A neural network provides better results over the existing approaches in speech recognition systems [1].

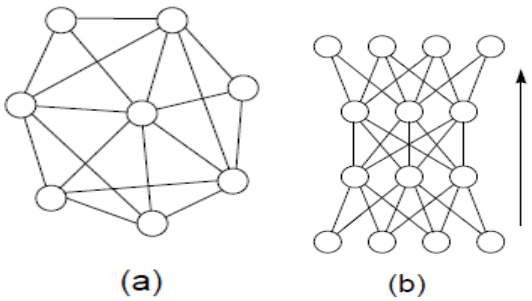
2. BASICS OF NEURAL NETWORKS
The basics of neural networks are discussed below. Neural Networks are of different types, but they all have four basic and common attributes:

- Processing units
- Connections
- Computing procedure
- Training procedure

.

2.1 Processing Units
A neural network contains several processing units, which are roughly analogous to neurons in the human brain. All these units activate in parallel and perform the task simultaneously. Processing units are responsible for the overall computation; there is no any other unit for the corporation of their activity. Each processing unit computes a scalar function of all its local inputs at every moment of time and then further broadcast the result to their neighboring units [2]. The units in a neural network are basically classified into input units, which used to receive data from the outside; hidden units, used to internally transform the data; and output units, which serve decisions or target signals.

2.2 Connections
All processing units in a neural network are organized in to a defined topology by a set of connections, or weights, shown as lines in a diagram. Each weight consist a real value, which ranging from - ∞ to + ∞. The value of a weight represents how much impact a unit has on its neighbor units weight with positive value represents excite the others by unit one, where as a negative weight represents that inhibits the other by unit one. The weights of all processing units are mainly one-directional, i.e. from input states towards output states, but it may be two-directional sometimes, especially when there is no any discrimination among input and output units.



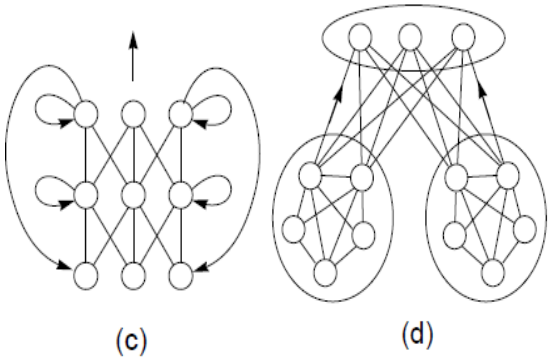


Fig 1: .Neural network topologies: (a) unstructured, (b) layered, (c) recurrent, (d) modular [3].

The above figure shows the topology of different neural networks. Unstructured Neural Network is mainly used in processor which is operated in parallel to provide the computational power for the speech recognition system. Layered neural net algorithms are self-organizing and their internal speech model maximizes the performance and produces better results comparable to existing systems. Recurrent Neural network are mainly used in pattern sequencing as it follows the sequences changes time to time. Modular neural network are used in designing of complex systems by using simple components.

Neural networks provide an approach for computation that copies the biological nervous systems. Algorithms implemented by using neural networks have been proposed to perform speech recognition tasks which people done with little possible effort.

2.3 Computation Procedure

Computation of neural networks begins after applying an input data to the input units of the network. Then the activation function of all the units is calculated either simultaneously or independently depends upon the structure of the neural network. The computation process in the unstructured network is termed as spreading activation and in layered network is known as forward propagation as it proceed from the input units to the output units. First we compute the net input of the processing unit and then we compute the output activation function of the net input to update a given processing unit.

2.4 Training Procedure

Training a network means enable the connections adaptive so that the network shows the actual computational behavior for all the input patterns. In training process usually weights are updated but sometimes the modification of network topology also takes place, i.e., addition and deletion of connections from the network topology [4]. Modification of weight is easy and beneficial than topology modification as a network with bulk connections able to set any of its weight zero, which is equivalent as deleting such weights.

3. RELATED WORK

The past research concluded that the use of the neural networks in the speech recognition system provides the better recognition result compared to the other existing approaches. The latest study of neural networks actually started in the 19th century, when neurobiologists first introduce the pervasive research of the human nervous system [5]. Cajal in 1892 find out that the nervous system is comprised of some basic unit's i.e. discrete neurons, which communicate with the other neurons through sending electrical signals down to their long *axons*, which finally activated and touch the *dendrites* of thousands of other processing units, which used to transmitting the electrical signals by *synapses*. Firstly, the different kinds of neurons were identified, and then analyze their electrical responses, and finally their patterns of connectivity and the brain's gross functional areas were mapped out. According to the neurobiologists study the functionality of individual neurons are quite simple and easy. Whereas to determine how neurons worked simultaneously to gain high level functioning, such as perception and cognition are very difficult.[6]

In 1943 McCulloch and Pitts proposed the first enumeration model of a neuron, named as binary *threshold unit*; output of that model was either 0 or 1 depends on whether its network input excel a given threshold value. There are various approaches proposed by the researchers to design an accurate speech recognition system for various purposes. In [7] Al-Alaoui algorithm is used to train the neural network. This method gives the comparable better results to the already implemented hidden markov model (HMM) for the recognition of the words. This algorithm also overcomes the disadvantages of the HMM in the recognition of sentences. An algorithm based on neural network classifier [8] for speech recognition used a new Viterbi net architecture which is recognized the input patterns and provided an accuracy of recognition rate more than 99% on a large speech database. This system is used for isolated word recognizer. In [9] author accomplishes the isolated word speech recognition using the neural network. The methodology of this approach is to extract the feature of speech signals using the Digital Signal Processing techniques and then classification using the Artificial Neural Network. This algorithm concluded that the better accurate recognition results are obtained from the probabilistic Neural Network PNN. In [10] author implemented a pre- trained deep neural network using the hidden markov model (DNN-HMM) hybrid architecture which is used to train the DNN to produce the better recognition results of large vocabulary speech database.

4. PROPOSED WORK

Speech recognition using the Epochwise Back propagation through time algorithm is proposed in this paper**.** In the proposed system neural network training is based on the calculation of epoch of the audio signal and then used these epoch value for the training of the neural network. The epoch values are basically instant of significant excitation of the vocal-tract system during production of speech. The input data sets used to train the neural network can be partitioned in to the independent epochs. Each epoch representing a temporal value of the input data. Back propagation neural network used in the system in following steps.

1. First we choose and fix the architecture for the network, which comprised of input, hidden and output units, all units will contain their sigmoid functions value.
2. Then we will assign the weights among all the processing units. The assignments of weights are random and usually between -0.5 and 0.5.
3. Each input pattern is used in order to re-train the weights in the neural network.
4. Next calculating each **epoch** for input audio data, a termination condition is checked.

In neural network architecture the weights of input and hidden layers are adjusted according to the target output values [11] The input data is considered as E which is proliferate through the topology so that we can easily note all the observed results $O_i(E)$ for the output units O_i. Meanwhile, we also note all the observed values $h_i(E)$ for the hidden units. After that, for each output unit O_k, we calculate its **epoch** as follows:

$$\delta_{o_k} = o_k(E)\big(1 - o_k(E)\big)\big(t_k(E) - o_k(E)\big) \qquad (1)$$

The epoch values from the output node are helpful to compute epoch value for the hidden nodes. This method is termed as Back Propagation because here we inseminate this data backwards by using the network [12]. For each hidden node H_k, we compute the **epoch** as follows:

$$\delta_{H_k} = h_k(E)(1 - h_k(E)) \sum_{i \in outputs} w_{k_i} \delta_{o_i} \qquad (2)$$

Here, we take the epoch value for each output node and then multiply it by the weight of the network from hidden H_k to the output nodes. After that we add all these simultaneously and multiply the summation by $h_k(E)*(1 - h_k(E))$. When we get all the calculated epoch values related with every unit (hidden and output), finally pass this information into the weight changes Δ_{ij} among units i and j. The calculation is defined as: for all weights w_{ij} between input node I_i and hidden node H_j, and summation of all units are as:

$$\Delta_{ij} = nH_j x_i \qquad (3)$$

Back Propagation Neural Network architecture is shown in below figure

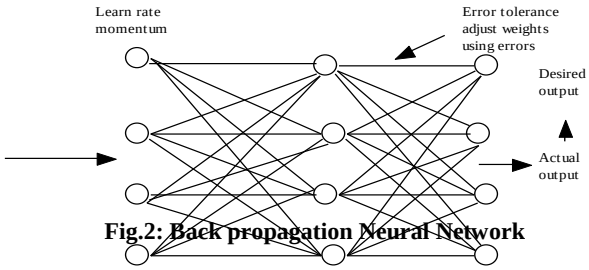


Fig.2: Back propagation Neural Network

The Back Propagation learning process requires the following components:

1. Set of input data patterns, input values and target output values.
2. Learning Rate Value.
3. Algorithm termination criteria.
4. Procedure for the updating weights.
5. Nonlinear functions or sigmoid functions.
6. Initial random weight values.

In proposed system Back propagation neural architecture contains the all above units and the basic steps of proposed system are defined as follows.

1. Read the input audio Signal.
2. Extract the epoch values
3. Train the neural network on the basis of epoch values.
4. Applied the back propagation neural network for the classification.
5. Matching the input data with the trained data.
6. Recognized the input.

Figure 2 shows the data flow diagram of the proposed system.

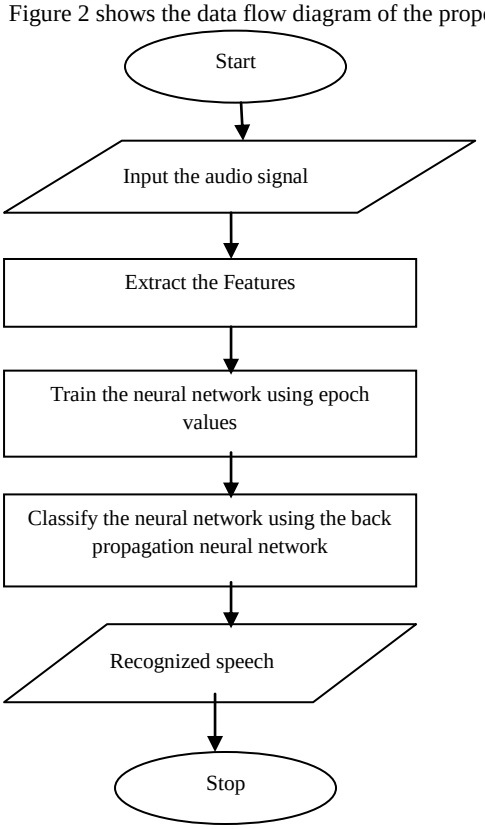


Fig.3: Proposed approaches for Speech Recognition

5. EXPERIMENTAL RESULTS

The experiment conducted on the audio database of English alphabets. Matlab R2010a (Math works) in Windows 7 was used to implement the proposed algorithm. The input signal is used to calculate the epoch values and then by using calculated epoch values the neural network is trained. The epoch rate of multi-layered neural networks topology based on training set could be computed by the number of unclassified data values. There are many output units, all of which produce wrong results (e.g., giving a value close to 1 when it should have output 0 and vice-versa), we have to be more conscious in our epoch value evaluation. In practice the overall network epoch is computed by using following formula:

$$\frac{1}{2} \sum_{k \in inputs} \Big(\sum_{k \in outputs} \big(t_k(E) - o_k(E) \big) 2 \Big) \qquad (4)$$

This calculation is not as complicated as it first shows up. The computation simply requires the working out the difference among the observed output for each output node and the target output and then squaring this to confirm it is positive, after that adding up all the squared differences for every output node and for every input signal.

Table. 1 Performance of Back propagation Neural Network

	Samples	MSE	% E
training	700	4.44301e-7	0
validation	150	9.55612e-7	0
Testing	150	5.84168e-7	0

Table 1 represented the performance of the back propagation neural network in terms of mean square error and percentage error. Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error. Percent Error indicates the fraction of samples which are misclassified. A value of 0 means no misclassifications, 100 indicates maximum misclassifications.

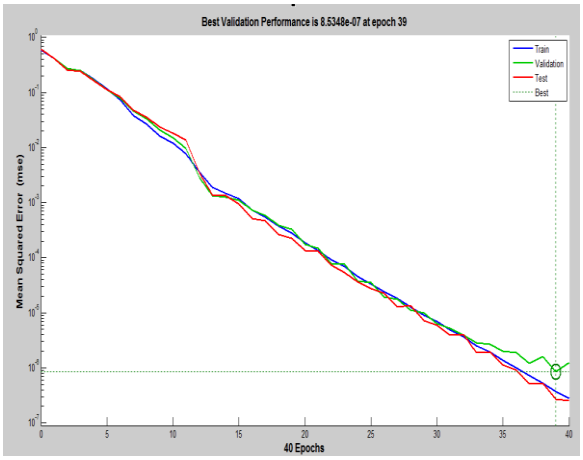


Fig.4: Performance graphs of epoch values.

The figure 4 shows the performance of the system based on epoch values. The best validation performance epoch value is selected from the different epoch on the basis of Mean Square error. The graph is plotted on epoch and means squared error values.

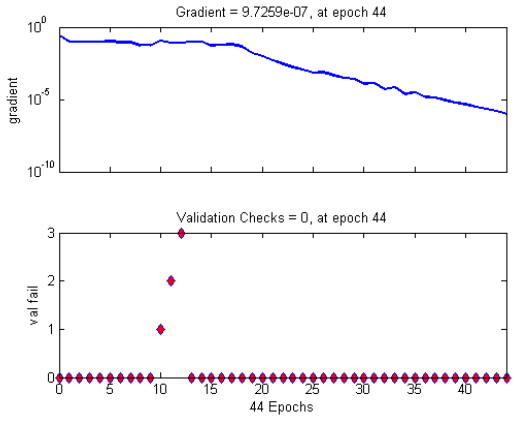


Fig.5: Error histogram of the training phase.

In the proposed approach the division of input data is random and the training of input samples is done on basis of scaled conjugate gradient. The performance is evaluated on the basis of the calculated mean square error.

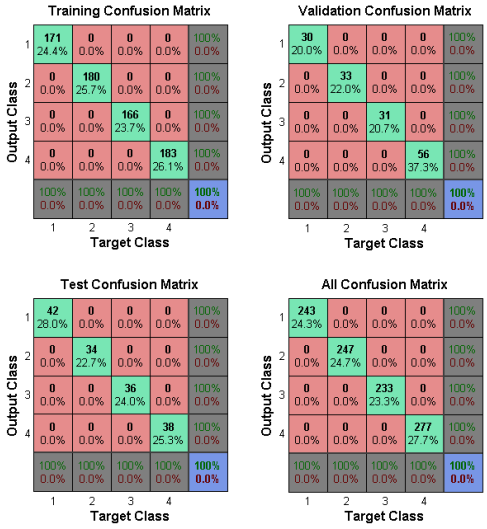


Fig.6: Confusion matrix of neural network.

Confusion matrix shows the classification of input samples in the various classes for the better visualization of the performance of the system.

Table: 2 Performance of the proposed system

Input alphabet	No of sound samples	Recognition rate
A	5	100%
B	5	100
C	5	100%
D	5	100%
E	5	98%
Total		99.6

In Table: 2 the recognition rate is calculated for five input English alphabet which is 99.6%. The new proposed Epochwise Back propagation through time algorithm yields the satisfactory results.

6. CONCLUSION

In this paper, we implemented the Epochwise Back propagation through the time varying epoch calculation. The experiment is conducted on the small set of English Language alphabet to calculate the recognition rate of the system. Some different sound samples (i.e., with different sampling frequency) of each alphabet are taken and used for testing the system. The above results show the performance of our proposed algorithm in speech recognition.

7. REFERENCES

[1] Jianliang Meng, Junwei Zhang,Haoquan Zhao, "Overview of the speech Recognition Technology", 2012 Fourth International Conference on Computational and Information Sciences.

[2] L. Fausset, Fundamentals of Neural Networks. PrenticeHall Inc., 1994, ch 4.

[3] Jiang Ming Hu, in the Yuan Baozong, Lin Biqin. Neural networks for speech recognition research and progress. Telecommunications Science, 1997, 13(7):1-6.

[4] H. Boulard and N. Morgan, "Continuous speech recognition by connectionist statistical methods," IEEE Trans. Neural Netw., vol. 4, no. 6, pp. 893–909, Nov. 1993.

[5] Mohamad Adnan Al-Alaoui, Lina Al-Kanj, Jimmy Azar, and Elias Yaacoub, "Speech recognition using Artificial Neural Network and Hidden Markov Model", IEEE Multidisciplinary Engineering Education Magazine Vol. 3, No.3, September 2008.

[6] RICHARD P. LIPPMANN, "Neural Network Classifiers for Speech Recognition" The Lincoln Laboratory Journal, Volume 1, Number 1 (1988)

[7] Gulin Dede, Murat Husnu Sazlı, "Speech recognition using artificial neural network." Digital signal processing © 2009 Elsevier Inc.

[8] George E. Dahl, Dong Yu, Li Deng, Alex Acero," Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition" , IEEE Transaction on Audio, and Language Processing, Vol. 20, No.1, January 2012.

[9] H. Paugam-Moisy, 'Parallel neural computing based on network duplicating',in Parallel Algorithms for Digital Image Processing, Computer Vision and Neural Networks, ed., I. Pitas, 305–340, JohnWiley, (1993).

[10] Stefano Scanzio, Sandro Cumani, Roberto Gemello, Franco Mana, P. Laface, "Parallel implementation of Artificial Neural Network Training for Speech Recognition." Pattern recogonition letters, © 2010 Elsevier B.V.

[11] N. Morgan and H. Bourlard, "Continuous speech recognition using multilayer perceptrons with hidden Markov models," in Proc. ICASSP,1990, pp. 413–416.

[12] Y. Hifny and S. Renals, "Speech recognition using augmented conditional random fields," IEEE Trans. Audio, Speech, Lang. Process., vol.17, no. 2, pp. 354–365, Feb. 2009.