International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

A COMPARATIVE STUDY OF ELMAN BACKPROPAGATION NEURAL NETWORK AND RADIAL BASIS FUNCTION NEURAL NETWORK MODELS FOR SPEECH RECOGNITION

Priyanka Tyagi
M. Tech student,
Department of Computer Science
Subharti University, Meerut
Uttar Pradesh

Dr. Jayant Shekhar
Professor (Director, SITE)
Department of Computer Science
Subharti University, Meerut
Uttar Pradesh

ABSTRACT - Speech is a most common way of human-to-human communication. The easiest and most common way of interaction is Speech; human-to-machine communication also becomes need. There is the technology to enable machine to understand process and recognize speech that is called Automatic speech recognition. ASR is the most fascinating area of pattern recognition. In this paper, we are analyzing the performances of Elman back-propagation neural network and Radial basis function neural network models for the recognition of speech signals. The work is organized in four stages: speech signal acquisition and pre-processing, training and test pattern vector creation, implementation & training of selected neural network models and comparative analysis of performances of selected neural networks.

Proposed work is conducted with 10 speech samples of English alphabets. To make them appropriate for further processing, digital signal processing operations are applied on signals to convert them. For training and testing of the network models, five feature pattern vectors are created. Performance of selected neutral network models is determine and examined for the created feature pattern vectors. Results indicate that elman back-propagation neural network model performs better than the Radial basis function neural network for all the test pattern vectors.

Keywords: ***Automatic speech recognition, Digital Signal processing, Sampling, Quantization, Elman backpropagation neural network, Radial basis network.***

I. INTRODUCTION

The most effective and the essential medium of communication between human beings is speech. It is also the means by which one can convey information about the personality, identity and linguistic information. In today's electronic era, speech has also become an important medium by which an human being can connect with a machine [1]. The popularity and acceptance of speech has attained such enormous heights that automatic speech recognition has become one of the most interesting area of pattern recognition. In the fields of engineering and science several techniques of speech recognition has been used in variety of application [2].

The technology by which a machine is made to learn and recognize the speech spoken by a person and converting it into the teed form is called automatic speech recognition technology. The characteristics of a person or the environment like speaking speed, noise elements etc. are the main features on which speech depends. The functioning of the speech recognition system is affected by these features. The working of a speech recognition system may be influenced by some others factors like varying speaking styles or accents age, emotional state of speakers ,etc.[3] The similarity of the modular structure of speech recognition system and the human mechanism to speech perception helps in handling such problems.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

Types of speech recognition systems:

1. ***ISOLATED WORD RECOGNITION SYSTEM-*** The isolated word recognition system accepts either a single word or single sound at a particular instance of time. These systems have only two states- listen and not-listen [4].

2. ***CONNECTED WORD RECOGNITION SYSTEM*** –These systems works by processing two or more sounds or words together. Words or sounds should have short silence between them [5].

3. ***CONTINUOUS SPEECH RECOGNITION SYSTEM*** –These systems work by accepting continuous speech between two sounds and there is no time boundary limitation on the input speech [6].

Till date, a number of techniques have been carried out so that the machines can recognize speech. Earlier, the focus was mainly based on designing machines which could speak rather than those which can understand and recognize speech. The Russian scientist Christian Kratzenstein, in 1782used resonance tubes connected to organ pipes and produced vowel sounds. Later, Charles Wheatstone, in 1879, produced various speech like sounds by building a speaking machine which had resonat or smadeofleather[7]. Since then, lot of researchers has been done in this area. The various efforts performed in the last decade are review and persistent meet.

Several sincere efforts and research works done in the area of automatic speech recognition using artificial neutral network model have been carried out, but till today, some gap still exists for optimal neural network architecture with high recognition accuracy and good generalized behavior.

In this paper, we are investigating the performance of Elman backpropagation neural network and Radial basis function network models for the speech recognition of alphabets of English language. An effort has also been made to present a comparative analysis of performances of both neural network models for noisy and noiseless input speech samples are also done.

This paper is further organized in five sections. Section 2 discusses the feature extraction process of the input speech samples. In section 3, implementation details of neural network models. Section 4 presents the simulation results, comparative study of recognition performances & accuracy of the selected neural network models and a complete discussion of the results. Section 5 considers the conclusion followed by references.

II. FEATURE EXTRACTION

Feature extraction phase is very important as the classification and recognition phase of a recognition process. Feature extraction is defined as the *"problem of extracting from the raw data the information which is most relevant for classification purposes, in the sense of minimizing the with-in class pattern variability while enhancing the between-class pattern variability"* [8].During the feature extraction phase, meaningful and significant features are extracted from a number of available features. These features should be independent of the orientation, location and location of the pattern for the smooth functioning of the recognition system. Thus, the goal of feature extraction process is to create an optimal feature vector to maximize the efficiency and to support classification process of the recognition process.
Data set used in the experiment consists of 10 speech signal samples English letters 'A', 'B', 'C', 'D' and 'E'. All speech signals are collected as audio files and each audio file is stored with the extension *.wav*. A set of 5input signals is presented in figure 1.

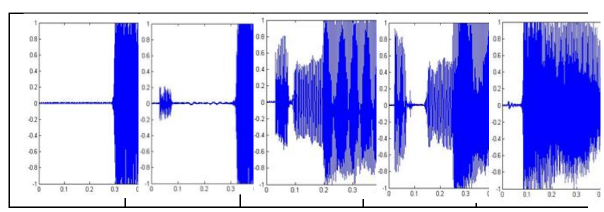


Figure1: A set of five input signal samples

Training pattern vectors are created by applying quantization and sampling operations on collected input signal samples. Sampling is used to convert the continuous-time analog input signal samples to the discrete-time signals as [9]:

$$y[m] = y(m \, . \, V_s) \qquad (1)$$

wherey[m]is the discrete time sequence signal
mis the sample index
Vis the sampling interval

The amplitude of y[m] obtained in this equation is known with infinite precision [10].To represent each

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

value an infinite number of digits are required instead of a finite number of digits. Quantization converts a continuous–amplitude signal into a discrete-amplitude signal. Thus, the amplitude of discrete-amplitude signal is known with finite precision. Result of the quantization process is the digital signal. Difference between the discrete-time signal and digital signal is called the quantization error [11].

In this paper, we create the test pattern vectors by introducing the 10%, 20%, 30%and 40% noise respectively to training pattern vectors of speech signals.

III. IMPLEMENTATION OF NEURAL NETWORK MODELS

An artificial neural network or ANN is a computational model which is designed to perform the complex pattern recognition tasks such as pattern classification, pattern mapping, pattern association, etc.[12] Thus, a neural network can be characterized as a computing architecture, which consists of a large number of simple highly interconnected data processing elements called neurons, designed to resemble the learning and storing capability of human brain for performing the task of pattern recognition [13][14]. These neurons work in parallel to learn& acquire knowledge and to make that knowledge available for use.
In this paper, we explore two neural network models trained with variants of gradient descent method of generalized delta learning rule. The basis of all deltas learning rule is back-propagation learning rule. Back-propagation (BP) is a supervised learning algorithm and belongs to a class of "learning with the teacher"[15]. Back-propagation is a systematic method of training artificial neural networks in which a predefined desired target output (t) for each input pattern is prepared. This target output is compared with the actual output (o) and difference is termed as error (E). The value of the error term is propagated backward from the output layer to hidden layer/s to update the weights in the hidden layer/s as [16]:

$$v_{kj} = v_{kj} + c\lambda(t_k - z_k)z_k(1 - z_k)o_i \qquad (2)$$

and the output layer as [20]:

$$w_{ji} = w_{ji} + c\lambda^2 o_i(1-o_i)x_i\left(\sum_{k=1}^{K}(t_k - z_k)z_k(1-z_k)v_{kj}\right) \qquad (3)$$

Mean of the error at the kth iteration is computed as:

$$E = E + \frac{1}{K}\sum_{k=1}^{K}(t_k - z_k)^2 \qquad (4)$$

Where w_{ji} : weights connecting input layer ith neuron to the jth neuron of the hidden layer,
w_{kj} : Weights connecting hidden layer jth to the kth neuron in the hidden layer,
λ : parameter used to control the gradient of the function,
t_k : output of the kth target vector and
o_i: output of the net output of the hidden layer neuron.

In the present work, the first neural network model we used is Elman backpropagation neural network.
Elman backpropagation network is a two layer backpropagation network in which a recurrent connection exists from the output of the hidden layer to its input [17] as shown in figure 2.

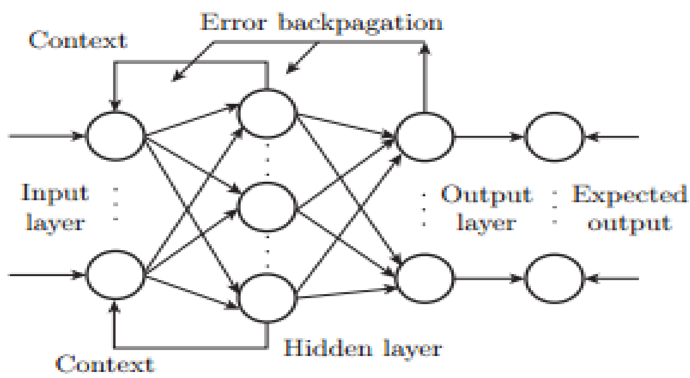


Figure 2: Elman backpropagation network

Network is trained with Gradient Descent backpropagation with adaptive learning rate. In this experiment, size of output vector T and input pattern vector P is same. The parameters used for the training and architecture of the network are presented in table 1.

Table 1: Parameters used for creating the Elman backpropagation neural network

Parameter	Value
Number of hidden layers	1
Number of neurons in hidden layer	10
Number of neurons in output layer	10
Transfer function for layer	Hyperbolic Tangent Sigmoid
Training function	Gradient Descent backpropagation
Maximum number of epochs	10000
Performance function	Mean squared error
Error goal	0.00001
Adaption rate	1.0
Back-propagation learning rate	0.1
Initial weights and biased term values	Values generated randomly between 0 and 1

Second neural network model in this work is Radial basis function network (RBF). RBF network is a three layer feed-forward neural network and consists of a single hidden layer in its structure, where hidden layer is non-linear and output layer is linear [18]. It's non-linear characteristics help to model the complex pattern mapping problems. In this network, the number of sample is more than the number of neurons in the first layer.

Gaussian function is widely used to compute the activation value for the unit of the middle layer. Gaussian activation function for the RBF network can expressed as:

$$z_j(X_l) = \exp\left[-\frac{\| X_l - \mu_j \|^2}{2\sigma_j^2}\right]$$

for j = 1,2,3,…………………, M (5)

where $\|.\|$: denotes Euclidean norm

x : N-dimensional input vector,

σ_j : width of the neuron, and

μ_j : mean of the j^{th} Gaussian function.

Two-fold learning is performed by updating the position and spread of centers as:

$$c_{ij}(t+1) = c_{ij}(t) - \eta \frac{\partial E}{\partial c_{ij}} \qquad (6)$$

for i=1 to size of the input pattern vector,

j=1 to hand up dating weights w to produce the desired output related to the input pattern vectors as:

$$w_i(t+1) = w_i(t) - \eta \frac{\partial E}{\partial w_i} \qquad (7)$$

where

η : learning parameter

The parameters used for the training and architecture of the Radial basis function network are presented in table 2.

Parameter	Value
Performance function	Mean squared error
Spread of Radial basis function	1.0
Number of neurons in layer 1	10
Number of neurons in layer 2	4800
Transfer function in layer 1	Radial basis transfer fun.
Transfer function in layer 2	Linear transfer function
Back-propagation learning rate	0.1

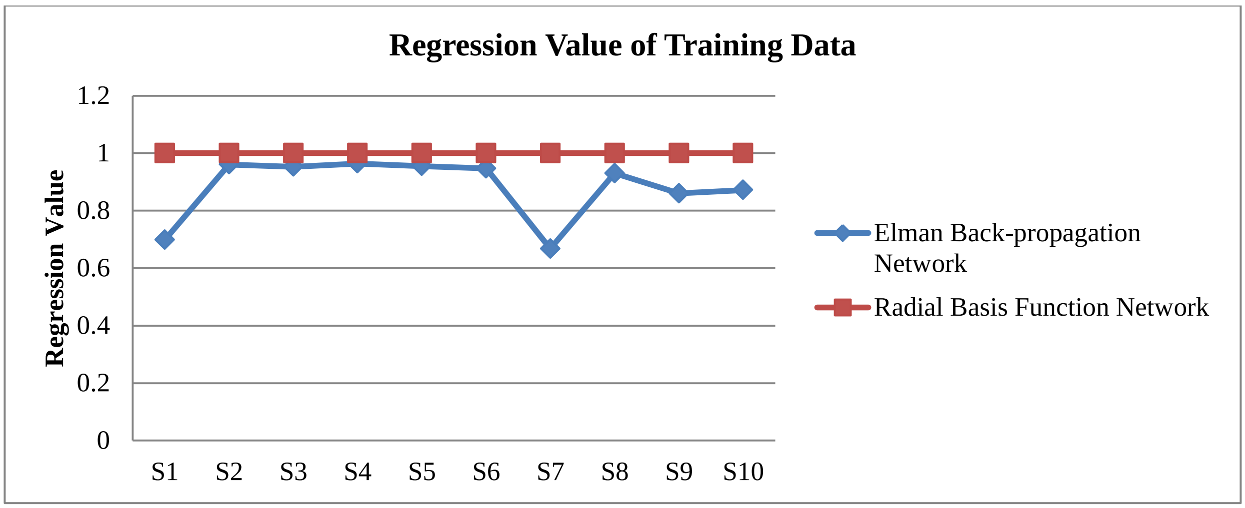
Table 2: Parameters used for creating the Radial basis function network

IV. RESULTS AND DISCUSSIONS

In the proposed simulation, we are analyzing the performance of elman backpropagation neural network and Radial basis function network models for created training pattern vectors and test pattern vectors of speech signals. The results presented in the simulation are considered from both selected elman backpropagation neural network models. Performance of these neural network models for the training patterns presented in table 5.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

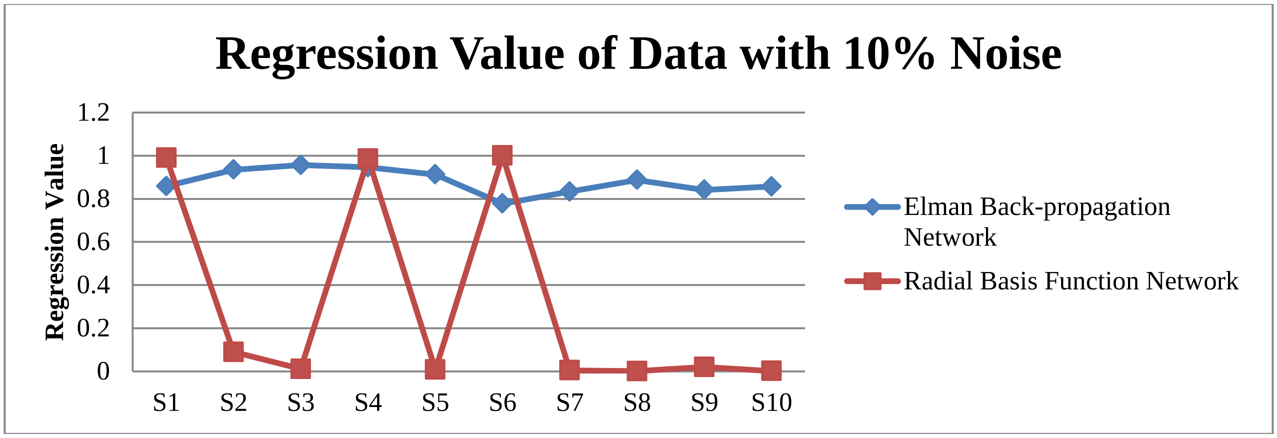
Network	Signals									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Elman Back-propagation Network	.6982	.9606	.9528	.9642	.9552	.9469	.6677	.9303	.8603	.8719
Radial Basis function Network	1	1	1	1	1	1	1	1	1	1



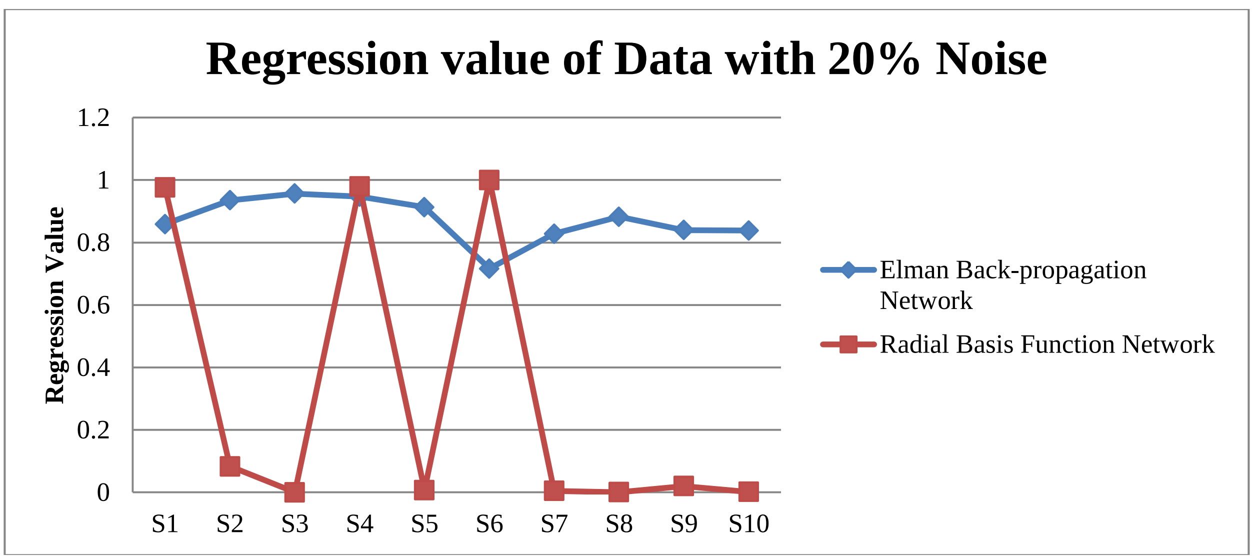
Speech Signal

Testing pattern vectors are created by introducing 10%, 20%, 30% and 40% noise in training pattern vectors of all signals. Performances of selected networks for testing patterns created for each signal are presented in table 2.

Network	Signals with 10% noise									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Elman Back-propagation Network	.8588	.9347	.9564	.9462	.9129	.7775	.8331	.8874	.8414	.8568
Radial Basis Function Network	.9904	.0898	.0111	.9865	.0072	1	.0045	.0008	.0198	.0017

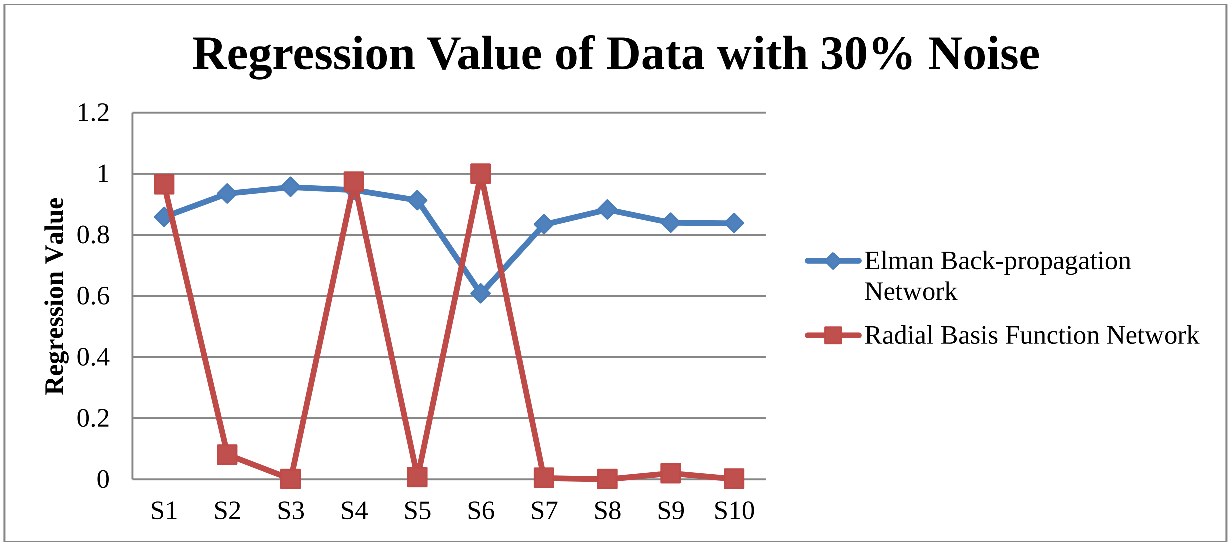


Network	Signals with 20% noise									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Elman Back-propagation Network	.8588	.9348	.9564	.9467	.9129	.71603	.8279	.8825	.8398	.8381
Radial Basis Function Network	.9766	.0829	.00002	.9795	.0072	1	.0045	.0008	.0198	.0017



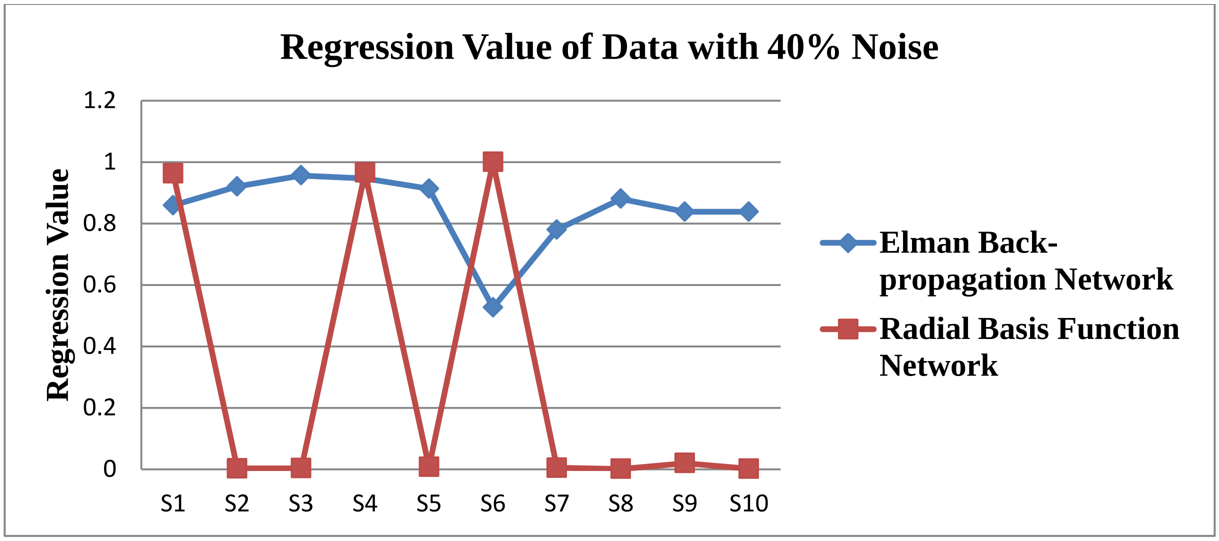
International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

Network	Signals with 30% noise									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Elman Back-propagation Network	.8588	.9348	.9564	.9467	.9129	.6080	.8339	.8825	.8399	.8381
Radial Basis Function Network	.9651	.0807	.0010	.9735	.0071	1	.0045	.0008	.0198	.0017



Network	Signals with 40% noise									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Elman Back-propagation Network	.8588	.92034	.9564	.9467	.9129	.5262	.7796	.880	.8381	.8381
Radial Basis Function Network	.9626	.0027	.003	.9654	.0071	1	.0045	.0008	.0198	.0017

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 64-72
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)



V. CONCLUSION

In this paper we analyzed the performance of Elman backpropagation network and Radial basis function network for the classification of speech signals of first five alphabets of the English language. Training pattern vectors are created by applying digital signal processing operations like quantization, sampling and coding to the input speech signals respectively. Test pattern vectors are created by adding 10%, 20%, 30% and 40% error respectively in the input signals used for training. Simulated results of the performance evaluation of the selected networks are presented and discussed. The following observations have been drawn from the simulated performance evaluation.

(i) The simulated results are also indicating that the Elman backpropagation neural network model shows above 67% recognition accuracy for training signal patterns. The networks shows similar recognition accuracy for test pattern vectors S1, S2, S3, S4 and S5. The network shows lowest recognition accuracy for signal S6, when the percentage of error increases.

(ii) The highest and lowest recognition accuracy presented by Elman backpropagation network is 96% for the signal S2and 53% for the signal S6 respectively.

(iii) The simulated results are also exhibiting that the Radial basis function neural network model shows 100% recognition accuracy for training pattern vectors and all test pattern vectors created for signal S6.

(iv) Results shows that Radial basis function network shows similar recognition accuracy for test pattern signals S5, S6, S7, S8, S9 and S10. Thus, the network is exhibit poor behavior of generalization for large data.

(v) Simulation results are showing that Elman backpropagation neural network exhibit good approximation and generalization.

(vi) Simulation results are showing that Radial basis function network shows a good approximation but poor generalization.

VI. REFERENCES

1. WouterGevaert, GeorgiTsenov and ValeriMladenov, "Neural Networks used for SpeechRecognition", Journal of Automatic Control, pg. 1-7, Vol. 20, 2010.
2. Gerasimos Potamianos, ChalapathyNeti, Guillaume Gravier, AshutoshGarg and Andrew W. Senior, "Recent Advances in the Automatic Recogntion of Audio-Visual speech', Proceedings of the IEEE, pp. 1-18, Vol. 91, No. 9, 2003.
3. NidhiSrivastava, "Speech Recognition Using Artificial Neural Network ", International Journal of Engineering Science and Innovative Technology (IJESIT), pg. 406-412, Vol. 3, Issue 3, 2014.
4. M.A.Anusuya and S. K. Katti, "Speech Recognition by Machine: A Review", Int. Journal of Computer Science and Information Security,pg. 181-205, Vol. 6, No. 3, 2009.
5. Santosh K. Gaikwad, Bharti W. Gawali and PravinYannawar, "A Review on Speech

Recognition Technique", International Journal of Computer Applications, pg. 16-24, Vol. 10, No. 3, 2010.

6. T. Landauer, C. Kamm, and S. Singhal, "Learning a Minimally Structured Backpropagation Network to Recognize Speech," In Proceedings of Ninth Annual Conference of Cogn. Sc.Soc., pp. 531–536, 1987.
7. Sir Charles Wheatstone, *The Scientific Papers of Sir Charles Wheatstone*, London: Taylorand Francis, 1879.
8. Shih F.Y., "Image Processing and Pattern Recognition - Fundamentals and Techniques", Wiley Pub. 2010.
9. John G. Proakis and Dimitris G. Manolakis, "Digital Signal Processing – Principles, Algorithms and Applications", Prentice Hall Publication, Third Edition, 2005.
10. DimitrisManolakis and Vinay Ingle, "Applied Digital Signal Processing – Theory and Practice", Cambridge University Press, First Edition, 2011.
11. Y. Tsividis, "Continuous-time Digital Signal Processing," *Electronics Letters*, pp. 1551–1552,Vol. 39, No. 21, 2003.
12. Yagnanarayana B., "Artificial Intelligence", Prentice Hall Pub., Ninth Edition, 2004.
13. Haykin S., "Neural Networks: A Comprehensive Foundation", New York: Macmillan College Publishing Company Inc., 1994.
14. Hecht-Nielsen R., "Theory of Backpropagation Neural Network", International Joint Conference on Neural Networks, pp. 593-605, Vol. 1, 1989.
15. Jesus O. D. and Hagan M. T., "Backpropagation Algorithms for a Broad Class of Dynamic Networks", IEEE Transactions on Neural Networks, pp. 14-27, Vol. 18, no. 1, 2007.
16. Hecht-Nielsen R., "Theory of Backpropagation Neural Network", Powell M.J.D., "Radial Basis Functions for Multivariate Interpolation: A Review", In Algorithms for the Approximation of Functions and Data, J.C. Mason and M.G. Cox, eds., Clarendon Press, pp. 143-167, 1987.
17. Gruning A., "Elman Backpropagation as Reinforcement for Simple Recurrent Networks", Neural Computation, IEEE Transcations, pg. 3108-3131, Vol. 19, Issue 11, 2007.
18. Chen S., Cowan C.F.N. and Grant P. M. "Orthogonal Least Square Learning Algorithm for Radial Basis Function Networks", IEEE Transactions on Neural Networks, pg. 302-309, Vol.2, No. 2, 1991.