IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sep 2018 E-ISSN: 2347-2693

Optimized Neural Network Architecture for The Classification of Voice Signals

Dipak D. Shudhalwar[1*], Ganesh Kumar Dixit[2], Pallavi Agrawal[3]

[1]Department of Engineering and Technology, PSSCIVE, NCERT, Bhopal, India
[2] Department of Computer Science, B. S. A. College, Mathura, India
[3]Department of Electronics and Communication MANIT, Bhopal, India

Available online at: www.ijcseonline.org

Accepted: 19/Sept/2018, Published: 30/Sept/2018

***Abstract*—** In this paper, the performance to optimize feed-forward neural network has been evaluated for the classification of voice signals of English alphabets. There are various feed forward neural network models have been used earlier but the selection of optimize architecture is a challenge. In this paper we are implementing a optimize architecture which is best suitable for the classification of voice signals. Digital signal processing operations are applied on analog speech signals to convert them into digital form and then to make them suitable for further processing by neural network models.

***Keywords*—** Digital signal processing, Optimize neural network, Pattern classification

I. Introduction

Voice recognition is an interesting and challenging task. In this task the inputs provided to the system may be highly variable [1]. Hence, to handle these problems, the modular structure of speech recognition system can be considered which is similar to the human mechanism of speech perception. It is considered that a speech recognition system can be an isolated word recognition system, a connected word recognition system, a continuous speech recognition system or a spontaneous speech recognition system [2].

It has been observed that the major problem with automatic speech recognition systems is to handle the variable length speech sequences. Hidden Markov Models (HMMs) are good in handling such data and modeling temporal behavior of the speech signals using a sequence of states. HMMs based phonemes recognition techniques were proposed in which feature pattern vector was created by extracting features from a single phoneme [3]. Further, HMMs and neural network models have also been used for speech recognition tasks [4, 5]. It is reported that the neural networks have been used extensively for many applications in speech processing and recognition such as speech synthesis, speaker adaptation and recognition, keyword spotting, etc. [6], [8].

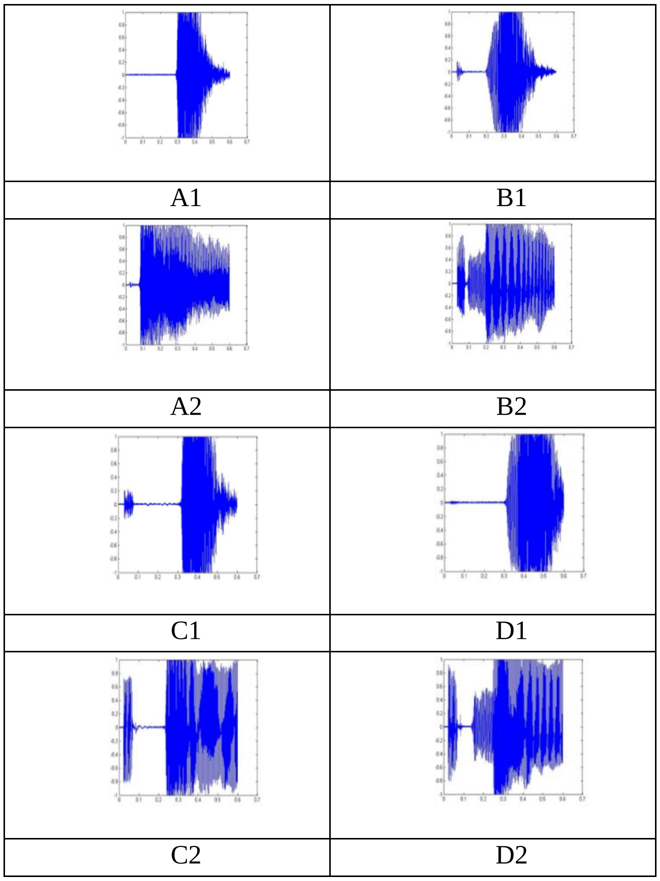
Backpropagation neural network with identification rate above 90% was used to perform and improve the accuracy of the classification of audio signals [9]. A multilayer perceptron neural network system optimized with genetic algorithm for classifying audio into speech and music was presented and wavelet transformation method has been applied for feature extraction [10]. It is reported that in this model system achieved 96.49%recognition accuracy.

Despite of the number of sincere efforts and research work done in the area of automatic speech recognition using artificial neural network, still there is some space left for the selection of optimal neural network architecture for recognition of speech with good accuracy. Therefore, in the present paper, we are investigating the performance of different feed-forward neural network models to select the optimal and suitable model for the speech recognition of first five alphabets of English language. Five feed-forward neural network models like Multilayer feed-forward network, Radial basis function network, Exact radial basis function network, Cascade feed-forward network and Elman back-propagation network have been selected for the experiment. A comparative analysis for the recognition accuracy of the selected neural network models for noiseless and noisy input speech samples is also conducted to explore the analysis of performances of the networks for the given speech samples.

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

II. FEATURE EXTRACTION

Ten (10) speech signal samples (two of each) of English letters ‘A’, ‘B’, ‘C’, ‘D’ and ‘E’spoken by a single female speaker are considered as input data samples. Individual alphabet is spoken in two different ways(quickly and slowly) for the time duration of 4 seconds each. All input speech signals are collected as audio files. The first five or quickly spoken speech signals are named as A1, B1, C1, D1 & E1; while the second set of speech signals, which are spoken slowly, are named as A2, B2, C2, D2 & E2. The set of collected input signals is presented in figure 1.



A1 B1

A2 B2

C1 D1

C2 D2

Figure1: Set of collected Input signal samples

These input signals are converted into digital form by applying two important digital signal processing operations, i.e. sampling and quantization, to make these speech input signals suitable for further processing of speech classification by neural network models. It has been widely accepted that most of the signals such as speech, radar signals and various communication signals are analog. Therefore, to process these analog signals it is necessary to convert them into digital form, i.e. to convert them into a sequence of numbers having finite precision.

These digitized signals in pattern vector form are used for training with feed-forward neural network models to generate the required signal classification. Further, to evaluate the performance of trained neural network models the test patterns are constructed. These test patterns are constructed by introducing 10%, 20%, 30%, 40%, 50%, 60% & 70% noise or error respectively to the signals used for training. These noisy analog signals are processed with sampling, quantization and coding steps for digitized presentation. These noisy digital signals are further presented as test pattern vectors and used to evaluate the performance of trained neural network model.

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. 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 [11], [12].

In this paper, an optimized feed-forward neural network model is explored. The neural network model which we have used in this work is Exact Radial basis function network (ERBF). We also have RBF network which is a three layer feed-forward neural network and consists of a single hidden layer in its structure, as shown in figure 3. In this architecture, the hidden layer is non-linear and output layer is linear. Hence, due to the non-linear characteristics, RBF is able to model the complex pattern mapping problems and exhibit the better generalization [13]. In this network, the number of neurons in the first layer is less than the number of samples and each unit implements a radial basis function such as Gaussian radial function, Quadratic function, Inverse quadratic function, Thin plate spline, etc. Hence, activation function of the hidden layer computes the Euclidean distance between the input vector and center of that unit and the value of the function increases or decreases monotonically with the distance from a center point.

Therefore we have used Exact Radial basis network which is also a radial basis network in which the basis function produces a network with zero error on training vectors. To make the network perform well, the spread is kept large enough so that at any given time point, the active input regions of the neurons have properly large output. The larger the spread is, the smoother the function approximation will be. This network does not perform well when the network is defined in terms of several input vectors. In this case the network produces as many hidden neurons as there are input vectors.

Therefore, for the optimization of the basis function, a number of techniques such as clustering algorithms, unsupervised learning, supervised learning, etc. have been proposed. But, to obtain the optimal performance, suprevised learning method is applied because it is required to include the target pattern vector in the training procedure. Thus, the weight vector modification and basis function parameters update is performed in iterative manner to accomplish the learning in supervised way. Hence, the update in weight and bais parameters at the mth step of iteration can be expressed as [14]:

$$w_{ij}(m) = w_{jk}(m-1) + \eta_1 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).\exp\left(-\frac{(\| x_i^l - \mu_{ki}^l \|)^2}{2\sigma_k^2}\right)$$

$$\mu_{ij}(m) = \mu_{jk}(m-1) + \eta_2 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).w_{jk}.\phi_k(x_i^l).\left(\frac{x_i^l - \mu_{ki}^l}{\sigma_k^2}\right)$$

$$\sigma_k(m) = \sigma_k(m-1) + \eta_3 \sum_{j=1}^{M}\sum_{k=1}^{K}(d_j^l - y_j^l)s_j^l(y_j^l).w_{jk}.\phi_k(x_i^l).\left(\frac{\| x_i^l - \mu_{ki}^l \|^2}{\sigma_k^2}\right)$$

Where η is the learning rate parameter, w_{ij} is the weight between the i^{th} unit of output layer and j^{th} radial unit of the middle layer.

IV. RESULTS AND DISCUSSIONS:

In the proposed simulation, the performance of an optimized feed forward neural network model has been analyzed for created training pattern vectors and test pattern vectors of speech signals. The results presented in the simulation are considered from the selected feed-forward multilayer neural network models. Performance of that neural network model for the training patterns are presented in table 5. Performances are presented on the basis of regression value after the complete training cycle.

Table 1: Regression value of training pattern vectors for all signals of the selected networks

Network	Signal				
ERBF network	A1	B1	C1	D1	E1
	1	1	1	1	1
	A2	B2	C2	D2	E2
	1	1	1	1	1

The performance of this network for the pattern vectors of input signals can also represent graphically in signal form. These signals are obtained as simulated output from the trained neural network models after presenting the digital input signals as shown in figure 2. Figure 2 is representing both the signals i.e., simulated output signal and actual input signal.

To analyze the performance of trained neural network for their generalized classification behavior the testing pattern vectors are created by introducing 10%, 20%, 30%, 40%, 50%, 60% and 70% error or noise in training pattern vectors of all signals. Performances of neural network for test pattern vectors are presented in table 2.

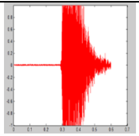
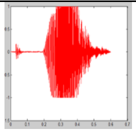
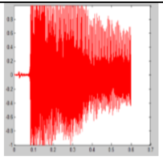
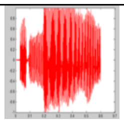
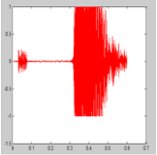
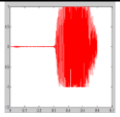
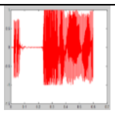
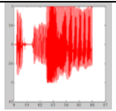
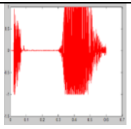
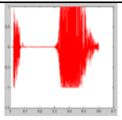
Signal	A1	B1
ERBF network		
Signal	A2	B2
ERBF network		
Signal	C1	D1
ERBF network		
Signal	C2	D2
ERBF network		
Signal	E1	E1
ERBF network		

Figure 2: Graphical representation of the performances of networks

Table 2: Signal-wise regression values of network for test pattern vectors

Error % → ↓ Voice signal	10%	20%	30%	40%	50%	60%	70%
A1	0.9968	0.9888	0.9856	0.0977	0.0547	0.0048	0.0048
B1	0.4496	0.4411	0.4350	0.1175	0.1175	0.1175	0.1175
C1	0.2718	0.0920	0.0904	0.0831	0.0830	0.0255	0.0255
D1	0.9945	0.9917	0.9889	0.9869	0.9830	0.0175	0.0175
E1	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235
A2	0.9313	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
B2	0.0130	0.0130	0.0130	0.0130	0.0130	0.0130	0.0130
C2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
D2	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304
E2	0.1929	0.1929	0.1929	0.1929	0.1929	0.1929	0.1929

Accuracy of recognition for test pattern vectors is taken as a measurement to analyze the performance of the neural network. Therefore, to do a comparative analysis for the performances of network model, the average of recognition accuracy is calculated for all signals with all chosen percentage of errors respectively as shown in table 3.

Hence, the performance analysis is considered by comparing the average of regression value between simulated output of the network performance for test pattern signal and expected output for the input signal. This comparison is presented in figure 3.

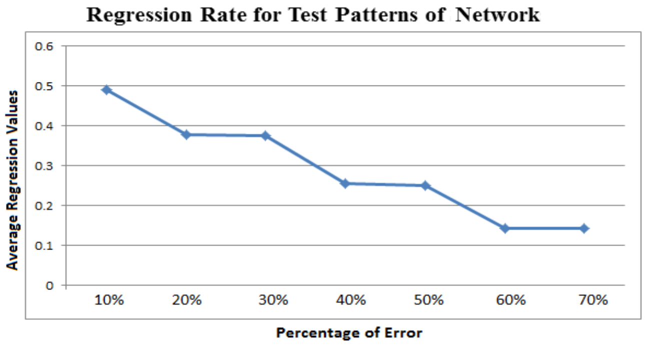


Figure3: Comparison of regression values of testing patterns for all selected neural network models

Tables 1 and 2 are exhibiting the comparison of performance of the ERBF Network for the training and testing pattern signals respectively. Results mentioned in table 5exhibit that, Exact radial basis function network model shows 100% approximation for all the training signal pattern vectors. The interesting results are obtained during the simulation test patterns, which show that the ERBF model shows better approximation for noisy test patterns. Most of the speech signals which are spoken quickly are better in comparison to spoken slowly and the performance degrades poorly when noise reaches level of 60% and above.

Results also display that Exact radial basis function network gives 100% recognition accuracy for all training patterns and all test pattern vectors of signal C2, while 90% or above recognition accuracy for few test patterns of signals A1, C1, D1and A2; but for most of the test pattern vectors, it shows reasonably low recognition accuracy. Thus, overall performances of exact radial basis function network are poor for test pattern vectors which have noise 60% or more.

V. CONCLUSION

In this paper we analyzed the performance optimized feed-forward neural network models trained with variants of back-propagation algorithm like Exact 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 sampling, quantization and coding to convert the continuous analog speech signal to discrete-time signal and digital signal respectively. Test pattern vectors are created by introducing 10%, 20%, 30%, 40%, 50%, 60% and 70% noise or 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.

Simulated results are showing that the performance of exact radial basis function network is better. Network model shows 100% recognition accuracy for training pattern vectors and all test pattern vectors created for the signal C2. The lowest test pattern recognition accuracy given by the system is .076% for signal A2.

REFERENCES

[1]. P. Rani, S. Kakkar and S. Rani, “Speech Recognition Using Neural Network”, In Proceedings of International Conference on Advancements in Engineering and Technology, International Journal of Computer Applications, pp. 11-14, 2015.

[2]. M. A. Anusuya and S. K. Katti, “Speech Recognition by Machine: A Review”, International Journal of Computer Science and Information Security, pg. 181 – 205, Vol. 6, No. 3, 2009.

[3]. X. Cui et.al., “A Study of Variable-Parameter Gaussian Mixture Hidden Markov Modeling for Noisy Speech Recognition”, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 15, No. 4, 2007.

[4]. G.E. Dahl, M. Ranzato, A. Mohamed and G.E. Hinton, “Phone Recognition with the Mean-covariance Restricted Boltzmann Machine”, Adv. Neural Inf. Process. Syst., No. 23, 2010.

[5]. D. Yu, L. Deng and G. Dahl, “Roles of Pre-training and Fine-tuning in Context-dependent DBN-HMMs for Real-world Speech Recognition”, In Proceedings of NIPS Workshop Deep Learn, Unsupervised Feature Learn, 2010.

[6]. H. Bourland and C.J. Wellekens, “Multilayer Perceptrons and Automatic Speech Recognition”, IEEE First International Conference on Neural Networks, San Diego, California IV-407-IV-416, June 21-24, 1987.

[7]. H. Yashwanth, H. Mahendrakar and S. David, “Automatic Speech Recognition Using Audio Visual Cues”, IEEE India Annual Conference, pp. 166-169, 2004.

[8]. Robinson and F. Fallside, “A Recurrent Error Propagation Network Speech Recognizer System”, Computer, Speech and Language, Vol. 5, No. 3, 1991.

[9]. L. Yang and Z. Yang, “Study on Audio Signal’s Classification Based on BP Neural Network”, IEEE Conference Publications on Artificial Intelligence, Management Science and Electronic Commerce, pp. 5153-5155, 2011.

[10]. S. Balochian, E. A. Seidabadand S. Z. Rad, “ Neural Network Optimization Genetic Algorithms for the Audio Classification to Speech and Music”, International Journal of Signal Processing, Image Processing and Pattern Recognition, pg. 47-54, Vol. 6, No. 3, 2013.

[11]. T. Lefteri H. and A. U. Robert, “Fuzzy and Neural Approaches in Engineering”, John Wiley and Sons Publications, 1997.

[12]. R. Hecht-Nielsen, “Theory of Backpropagation Neural Network”, International Joint Conference on Neural Networks, pp. 593-605, Vol. 1, 1989.

[13]. M.J.D. Powell, "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.

[14]. S. N. Parappa and M. P. Singh, "Conjugate Descent of Gradient Descent Radial Basis Function for Generalization of Feed-forward Neural Network", International Journal of Advancements in Research & Technology, pg. 112-125, Vol. 2, Issue 12, 2013.

Authors Profile

Dipak D. Shudhalwar, M. Sc., M. Phil., Ph. D. in Computer Science, is Associate Professor (CSE) and Head, Department of Engineering and Technology, PSSCIVE, NCERT, Bhopal. He is having more than 22 years of experience Research, Development and Training. He is basically working for the design and development curricula and instructional material for the various vocational courses in IT, Electronics, Telecommunication, Media and Entertainment under NSQF at NCERT, Bhopal. His research area includes Software Engineering, Component Based Software Reliability, Artificial Neural Network and Soft Computing. He has more than been 10 papers published in the reputed journals and several in the conferences. He has developed more than 30 curricula, 30 textbooks and organised more than 100 training programmes at NCERT.

Dr. Ganesh Kumar Dixit, M.C.A., Ph. D. in Computer Application, is Assistant Professor in Computer Science, B. S. A. College, Mathura, U.P. He is having more than 15 years of teaching experience in the various subjects of Computer Science. He is also a working group member for design and development curricula and instructional material for the various vocational courses in IT under NSQF at PSSCIVE, NCERT, Bhopal. His research area includes Pattern Recognition, Artificial Neural Network and Soft Computing. He has more than been 5 papers published in the reputed journals and several in the conferences.

Pallavi Agrawal has received her BE(Hons.) in Electronics and Communication Engineering from RGPV, Bhopal, INDIA and MTech in Digital Communications from Maulana Azad National Institute of Technology (MANIT), Bhopal, INDIA. She is now a PhD scholar in Electronics and Communication Department at MANIT Bhopal, INDIA, under the supervision of Dr. Madhu Shandilya. Her areas of research and interest are in the field of Digital Speech Signal Processing, Digital Communication and Statistical Signal Processing and Artificial Neural Networks.