Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 2, Issue 19; October-December, 2015, pp. 1680-1683

http://www.krishisanskriti.org/Publication.html

Digit Recognition by Supervised Learning using ANN

Ramen Barman

Deptt. Of Computer Science, B.B.K. College, Nagaon, Barpeta, Assam, India, Pin- 781311.
E-mail: ramenbarman89@gmail.com

Abstract—*Through the present study an attempt is made to verify and recognize the speech by person through computing system. At first, an artificial neural network is designed i.e. simulation of human brain is done with the networking facilities of neuron. Then that neural network is trained up with different types of speech of different people. Then testing purpose is made to verify and recognize them.*
The main aim of this research is to focus on offline recognition process. Some set of sound of different persons was collected and stored for verification process. These sounds are used to train the neural network. This offline speech recognition system is simple and do not require any additional tools.

It Introduced Artificial Intelligence in providing an approach and solutions on Digit recognition problems, it also improving the perspective and awareness of people about the importance of artificial intelligence in their everyday life.

Keywords*: Neural Network, Speech Recognition, Artificial Intelligence, Digit Recognition.*

1. INTRODUCTION

The purpose of speech is communication. Speech recognition is a method that uses an audio input for data entry to a computer or a digital system in place of a keyboard. In simple terms it can be a data entry process carried out by speaking into a microphone so that the system is able to interpret the meaning out of it for further modification, processing or storage [1].

Voice (or vocalization) is the sound generated by humans, animals and other vertebrates using a combination of the lungs and the vocal folds in the larynx, or voice box. Speech is a bit different. Speech contains emotions and feelings and is generated by precisely coordinated muscle actions in the head, neck, chest, and abdomen. Speech results after a gradual process involving years of learning and practice [2]. Keyboard, although a popular medium is not very convenient as it requires a certain amount of skill for effective usage. A mouse on the other hand requires a good hand-eye co-ordination. It is also cumbersome for entering non-trivial amount of text data and hence requires use of an additional media such as keyboard. Physically challenged people find computers difficult to use. Partially blind people find reading from a monitor difficult. Current computer interfaces also issue me a certain level of literacy from the user. It also expects the user to have certain level of proficiency in English. In our country where the literacy level is as low as 50% in some states, if information technology has to reach the grass root level, these constraints have to be eliminated. Speech interface can help us tackle these problems.

Speech Synthesis and Speech Recognition together form a speech interface. A speech Synthesizer converts text into speech. Thus it can read out the textual contents from the screen. Speech recognizer had the ability to understand the spoken words and convert it into text. We would need such software's to be present for Indian languages [3].

2. AREA UNDER INVESTIGATION

The present study is an attempt to verify and recognize assamese digit uttered by different speaker. It is try to develop a system by collecting some voice sample of different speaker to investigate the present study.

3. EXPERIMENTAL METHOD:

3.1: Data Acquisition

A sound database of 100 recorded sound is collected from 10 different speker, each speker recorded 10 assamese digit from 0 to 9 . All sounds are recorded using gold wave sound recorder at a sampling rate of 22050 at 16 bit mono channel.

3.2: DESIGN OF NEURAL NETWORK

3.2.1: MLP Classifier: The multilayer perception (MLP) neural networks can be used as a classifier and usually perform well even when there are a few training samples. Since their capability in nonlinear estimation and accuracy in interpolation, MLPs are known as a high performance classification in speech recognition systems.

On the other hand, after MLP learning, MLP weights can be updated during working as a classifier based on valid patterns and so, can learn slow changes of speech prototype that is very important in speech recognition.

The MLP with one hidden layer and sufficient nodes in this layer can estimate every nonlinear function. The proposed MLP for speech classification has three layers. The number of nodes in first layer should be equal to the number of features in the signal features string (2×Ns) and the nodes in output layer must be equal to the number of signal classes. The numbers of nodes in the hidden layer are related to Ns, number of classes and complexity of signal. [4]

Another advantage of using MLP as a classifier is the controlling of classification rejection. When applying a proper threshold on output of the MLP, the rejection of classifier can be simply controlled.

Based on the number and complexity of the signals, Ns that are the number of re-sampled points of signal and number of nodes in hidden layer of MLP are selected.

3.2.2: Network Parameters

The following discusses the different network parameters and architectural issues explored during experimentation, followed by an examination of MLPs with two hidden layers and the different training scenarios.

- Number of Nodes in the Hidden Layer: This is the most influential adjustable parameter within the model constraints. Architecture determination can be treated as an optimization problem exploring various possible designs looking for the most suitable structure. In a MLP with one hidden layer, once the training features are decided upon, the optimization problem reduces to a decision over the number of units in the hidden layer. This is a linear search of a noisy function (each time a network is trained a slightly different error rate results). [5] In the NN SRS system, the search involved imposing a minimum of two and a maximum of 120 (roughly three times the number of input units) on the number of hidden units and experimenting exhaustively within this range. Values between fifteen and thirty proved to be the most successful (in terms of OER) with nineteen hidden nodes used in any further NN experimentation.[6]
- Learning Rates: The NN learning rate controls the speed with which the network learns. A higher learning rate causes the algorithm to converge faster but may introduce instabilities, especially if the data is noisy. Experimentation involved exhaustively searching all learning rates from 0.05 to 0.95 with increments of 0.5. Small changes in the learning rate did not have a significant influence on the final error rates and a value of 0.6 produced consistently acceptable results. Degradation occurred when the learning rate moved closer to zero or one.
- Momentum: Small changes in the momentum value did not have a large impact on system accuracy. A value of 0.3 was used in the final system.
- Activation Functions: Sigmoidal activation functions were used in each of the NN nodes as they are generally more appropriate for HSV applications.

3.3: TRAINING AND TESTING

3.3.1: Training: The second stage of MLP learning is training during the system working that is necessary due to small changes in human speech. In our proposed scheme, when system recognize the sound of a digit and if the capturing time between the first stored sound and the current one is more than a threshold , the oldest one is replaced by the newest. For adding the new sounds to its trained set, it is necessary to train MLP by a few epochs. In this case, all of stored sounds are used as training set [7].

- We suggest that regarding the gradual changes in a given sound during a limited time interval, adding the newer instances of each sound to the training set and re-training of the network will improve the classification performance. Also, considering the characteristics of gradient based training algorithms and the interpolation ability of the artificial neural networks, this re-training compensates the slight input changing.
- On the other hand, if a new sound prototype is asked to be added to the recognition system, MLP architecture can be changed and the learning process with new samples as well as previous ones will upgrade the new classifier. In this case, system needs to have sufficient learning epochs to reach to an acceptable error. To determine the rejection threshold, a few unknown speeches (which were not in the data set) applied to the system then the largest output value was considered as the rejection threshold [8].

3.3.2: Testing: We need to test our network to see if it has found a good balance between memorization (accuracy) and generalization**.**

3.4: Neural Networks in Matlab:

- Matlab has a suite of programs designed to build neural networks (the Neural Networks Toolbox). Additionally, there are demonstrations available through Matlab’s help feature. In this lab , we will only work with three layer “feed forward” nets .
- There are three steps to using neural networks: Design, Training, and testing .

Design

- Define the number of nodes in each of the three layers (input, hidden, output).
- Define what is for each of the nodes (usually all nodes in a layer will all use the same). These are called the transfer functions.

- Tell Matlab what optimization (or training) routine to use. Generally, we will use either traingdx, which is gradient descent, or trainlm (for Levenburg-Marquardt, which is a combination of gradient descent and Newton's Method).
- Optional parts of the design: Error function (Mean Square Error is the default), plot the progress of training, etc.

Training

- Once the network has been designed, we "train" the network (by optimizing the error function). This process determines the "best" set of weights and biases for our data set [8].

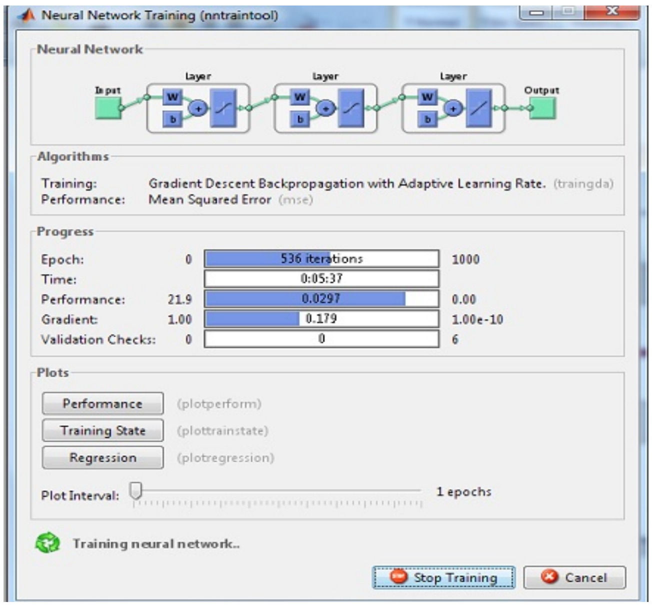

Fig. 1: Matlab Neural Network Training.

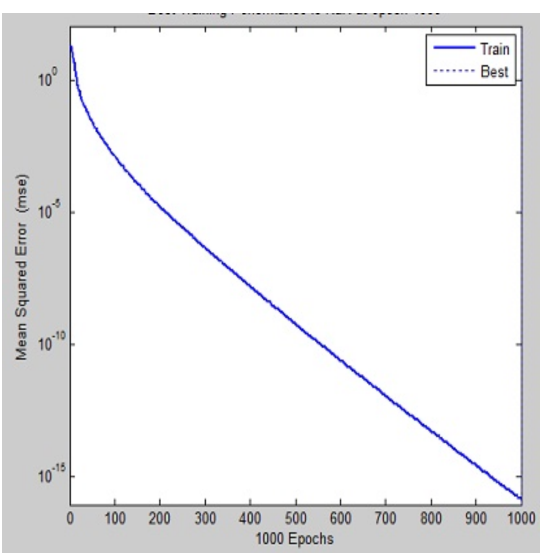

Fig. 2: Training Performance

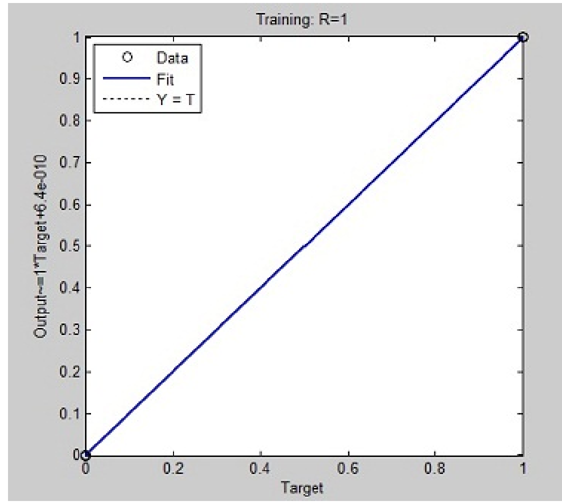

Fig. 3: Training Regression

4. RESULTS AND DISCUSSIONS

4.1 Testing and results

The sound signals are preprocessed. The distances are then fed to the Column Vector Conversion module and inputs for the corresponding neural networks are formed in the same way as in the training stage.

The networks are simulated with the corresponding inputs. Classification is done on the basis of outputs of simulation of all the networks at a time. Threshold condition is set to take care of the tradeoff between FAR and FRR.

Algorithm

Given a test Sound for recognition. The objectives are:

i. The database sounds are considered and the neural network is trained by each sound.

ii. The test sound features are compared with the database sound features using NN.

Results

i. The sounds given for the recognition are recognized and display the image and play the sound at the same time.

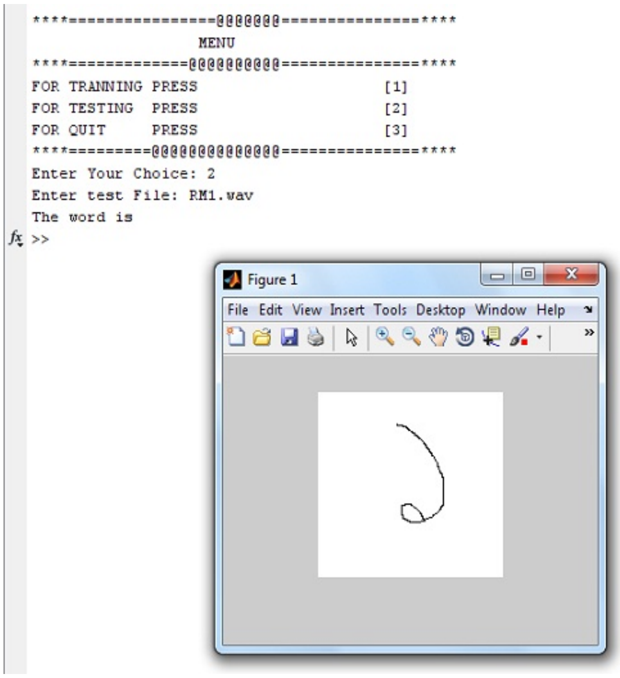

Fig. 4: Test Result for the wav file RM1.wav.

5. CONCLUSION

In most speech recognition tasks, human subjects produce one to two orders of magnitude less errors than machines. There is now increasing interest in finding ways to bridge such a performance gap.

The model is designed to recognized the assamese digit 0-9(sunya to na). For these purpose the system is trained with 100 sound files from ten different speaker, from which 5 are male and 5 are female. I used Matlab for Design, training and testing; the trained model was used to recognize the trained sound. Recognizer gave good results when tested for sound used for training the model and can be used for large data set too after trained the data sets. For other sounds the result is not worked. Future systems need to have an efficient way of representing, storing, and retrieving knowledge required for natural conversation.

REFERENCES

[1] The free encyclopedia Wikipedia, Speech.

[2] M.A.Anusuya and S.K.Katti, Department of Computer Science and Engineering, Sri Jaya chamarajendra College of Engineering Mysore, India “Speech Recognition by Machine: A Review”.

[3] Ripul Gupta, “Speech Recognition for Hindi” Department of Computer Science and Engineering Indian Institute of Technology, Bombay, Mumbai.

[4] L.R.Rabiner and Ronald W. Schafer, "digital processing of speech signals", Pearson Education, 1997.

[5] Ocw.mit.edu/NR/rdonlyres/Electrical...Speech.../lecture2.pdf, Acoustic Theory of Speech Production.

[6] Stephen Cook, "speech recognition how to, stephen cook,”

[7] Ajith Abraham, Oklahoma State University, Stillwater, OK, USA, Artificial Neural Networks.

[8] Neural Networks in Matlab, neural Lab1.pdf

Serdang, Selangor Darul Ehsan, Malaysia,” DIGIT RECOGNITION USING NEURAL NETWORKS”