Comparison of BFO and Back-Propagation Neural Network for Isolated Word Recognition

Pushpa Rani
Department of Computer Science & Engineering
Chandigarh University
Research Scholar

Arvinder Kaur
Department of Computer Science & Engineering
Chandigarh University
Assistant Professor

ABSTRACT
This paper provides comparison between BPNN and BFO for isolated word recognition in English language. In this paper, eleven English words were recorded from ten speakers including both male and female and have been recognized. The features of these spoken words were extracted using Mel Frequency Cepstral coefficient algorithm. Classification is done using back propagation neural network (BPNN) and bacterial foraging optimization algorithm (BFO). In an output we get meaning of that English spoken word in Hindi. This Hindi meaning is also a voice sample. Thus our input is a voice sample and our output is also a voice sample. All this implementation is carried out in Matlab platform. The current research work has successfully compared two algorithms on the basis of their performance namely BPNN and BFO. The research work has analyzed that BFO provides a better accuracy, varying from 15 to 20% more accurate than BPNN.

Keywords
Bacterial Foraging optimization, MFCC, Neural networks, BPNN, Speech recognition.

1. INTRODUCTION
Word recognition in the world of speech processing is one of the major concerns for researchers in today's world. As different countries have different accent of speaking. Indian accent is quite different from the others. The pitch of Indian accent is flat and hence it is often seen that recognition devices do not provide 100% accuracy for the spoken word. Speech recognition is one of the most critical topics in natural language processing. It provides interactive human-computer interaction. Computers, security systems and cellular telephones are some of it real time applications [1]. In spite of good progress, speech recognition process is still facing lots of problems. It is affected by the background noise while recording the voice samples. If the same word is pronounced differently by different speakers then it becomes very much difficult to identify each of them because of their age, accent, sex, emotional condition etc [2].

Voice recognition involves two selections for the classification namely training and testing [1]. In training phase voice samples are recorded from different speakers then their features are extracted using various techniques like MFCC, LPC and stored in the database. In testing phase pattern matching is done using various techniques like DTW, HMM, SVM, ANN, BPNN [3] [6] [7]. The current problem of this research work is to compare the work of Back propagation neural network (BPNN) and another optimization algorithm of this series known as bacterial foraging optimization (BFO). The word which has been taken in contrast are accident, signature, bangle, bucket, bright, clock, direction, literature, trouser, shoes and mountain. The problem statement also involves formation of input and target set with the help of MFCC algorithm. The extracted feature would be passed to both BPNN and BFO for classification and optimization. The parameter of evaluation would be accuracy for each word recognized.

The main disadvantage of back propagation is its convergence rate is relatively slow and it is often trapped in local minima [14]. So to solve this problem bacterial foraging optimization technique is used to enhance the learning process in terms of convergence rate and classification accuracy.

2. BASIC STRUCTURE OF SPEECH RECOGNITION
Speech recognition process is divided into many different modules illustrated in Figure.1. First of all voice samples are recorded using microphone in a noise free environment. Then filter the voice samples using Praat tool. Features are extracted using Mel frequency cepstral coefficients (MFCC) technique. All these coefficients are stored in the database.

In recognition mode voice sample is captured and its features are extracted using MFCC. Further classification is done that relates the input sound to the best fitting sound in a known dataset.

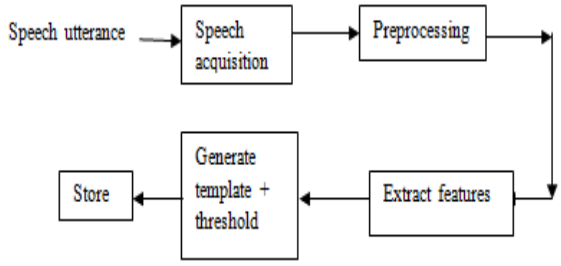

Figure: 1. Training mode

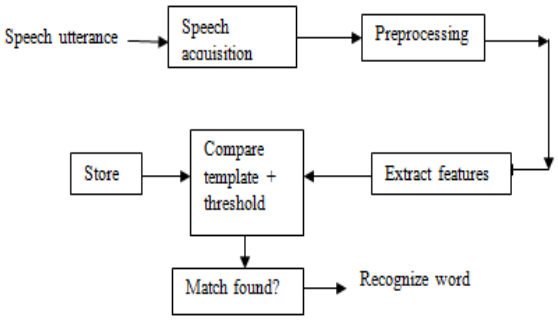

Figure: 2. Recognition mode

3. FEATURE EXTRACTION

MFCC is the most powerful method of feature extraction. It is designed using the knowledge of human auditory system. Voice samples are recorded in a noise free environment. Segmentation of voice sample into various frames. Each frame is further multiplied with hamming window to reduce discontinuity. FFT is applied to convert time domain into frequency domain. The output of FFT is a spectrum. Then mel-bank filter is applied so as to get smooth magnitude spectrum.

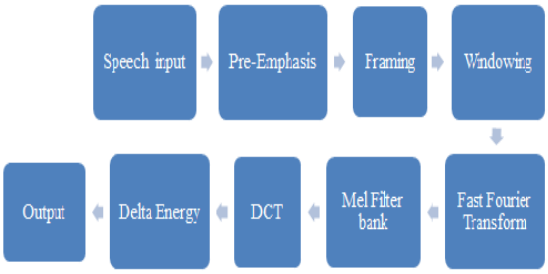

Figure: 3. MFCC block diagram

DCT is applied to get mel-scale cepstral coefficients illustrated in Figure 4 and Figure 5. Energy can also be calculated, it can be another parameter. Acceleration and velocity can also be calculated. The number of obtained coefficients is 39. The performance of MFCC is improved by adding log energy or delta function. Plot of 14 MFCC coefficients is shown in Figure.4.

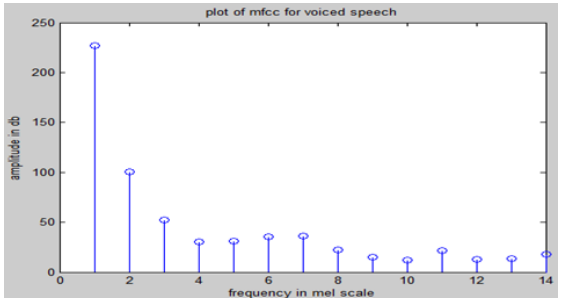

Figure: 4. Plot of MFCC coefficients

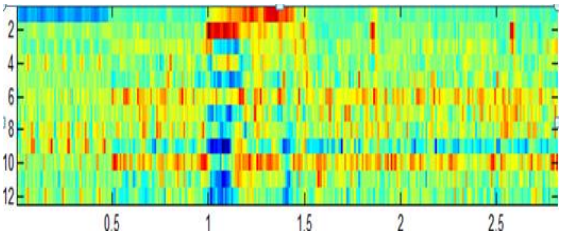

Figure: 5. Plot of Mel frequency cepstrum

4. CLASSIFICATION

4.1 Back Propagation Neural Networks

Neural networks are used for classification, function approximations or mapping problems that are tolerant to some indistinctness, have huge training set and where no hard and fast rules can be applied. Because ANN closely approximate unknown functions to desired accuracy. Basic functions of ANN are learning and adaptation. ANN learns from its surroundings and simultaneously improves the performance of the model. ANN consists of three layers that are input layer, hidden layer and output layer. During the learning process weights of the neural networks are optimized until a certain criteria is met. To train the neural network for supervised learning back propagation algorithm is used. It is also known as generalized delta rule. Input is given to the network connecting weights during the training period. Initially random values are assigned to weights and the output is obtained. This process is repeated again and again, modifying the weights until the network error reduces to zero. A single hidden layer is sufficient for many applications as more hidden layers make training slower.

$$\text{Error}=\frac{1}{2}\sum\sum(\text{Target output-Actual output})^{2}$$

Back propagation neural network is used to solve many real world problems but the main disadvantages of this algorithms are poor local optimal convergence and poor performance.

So bacterial foraging optimization algorithm can be used to enhance the learning process in terms of convergence rate and classification accuracy.

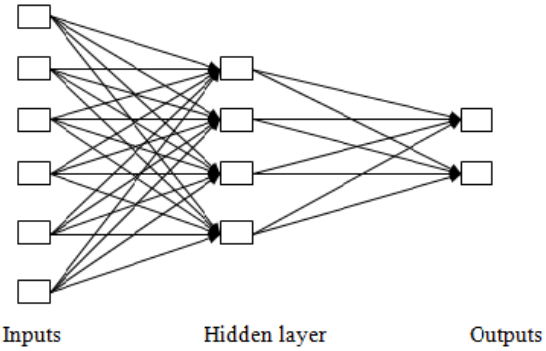

Figure: 6. Simple Architecture of ANN

4.2 Bacterial Foraging Optimization Algorithm

Passino proposed a new algorithm for distributed optimization in 2002. This method depends on the foraging behavior of the bacteria named as E coli found in human intestine. Animals having poor foraging strategies are reduced and support the propagation of those having rich foraging strategies. Process is Explained below:

1) Chemotaxis: Normally a bacterium attracts towards the nutrients. Bacteria moves basically by swimming or tumbling.

$$\theta_i^{t+1}=\theta_i^t + C(i)\,\phi\,(\text{i}) \quad \text{..... eq. 1}$$

$$\phi\,(\text{i}) =\frac{\Delta_i}{\sqrt{\Delta_i \Delta_i^t}} \quad \text{..... eq.2}$$

θ_i^t is the present position of the *i*th bacterium in ith chemotaxis step. Δ_i is the random vector.

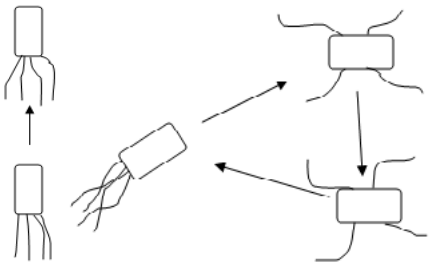

Figure: 7. Swim and tumble of bacteria

2) Reproduction: After obtaining sufficient food bacterium in the population will reproduce. Some of the bacteria will die also but the size of population remains same.
3) Eliminate and dispersal: Due to changes in the environment some of the bacteria will die and are eliminated. Thus it optimized the number of bacteria. Resulted bacteria are the best one and healthy.

The current problem of this research work is to compare the work of both BPNN and BFO. The word which has been taken in contrast are accident, signature, bangle, bucket, bright, clock, direction, literature, trouser, shoes and mountain. The problem statement also involves formation of input and target set with the help of MFCC algorithm. The extracted feature would be passed to both BPNN and BFO for classification and optimization. The parameter of evaluation would be accuracy for each word recognized. It is found that bacterial foraging optimization increases the performance of the system and is much better than BPNN.

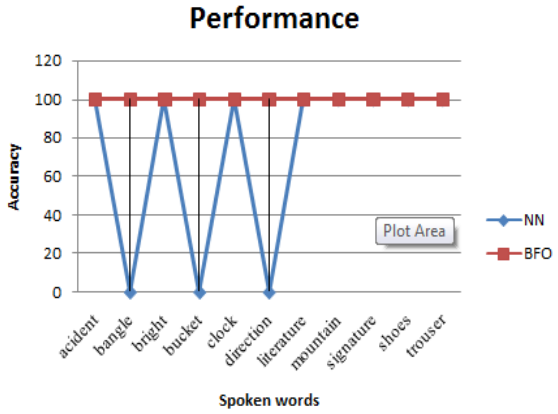

Figure: 8. Performance evaluation graph

Table1: Comparing the result for BPNN and BFO

Serial no.	Words	BPNN	BFO
1	Accident	100	100
2	Bangle	0	100
3	Bright	100	100
4	bucket	0	100
5	clock	100	100
6	direction	0	100
7	literature	100	100
8	mountain	100	100
9	shoes	100	100
10	signature	100	100
11	trouser	100	100

5. CONCLUSION

The neural network is a very powerful speech signal classifier. The pre-processing quality is giving the biggest impact on the neural network performance. Mel Frequency cepstral coefficient is a very reliable tool for pre-processing stage. MFCC is better than other techniques since logarithmic frequency decomposition mimics the human auditory perception system better than uniform frequency decomposition. MFCC has the ability to represent the speech amplitude spectrum in a compact form. The Cepstral coefficients obtained are the feature vectors that are stored in a .mat file and are given to the back-propagation neural network or BFO for classification.

With the present research work, it was concluded that optimizing the feature set would result into better classification than passing the value directly to the classifier. In this research work the comparative study of optimization and classification method has been provided and it has been seen that once the data is optimized it provides better classification accuracy rate.

The current research work provides an optimistic way of matching using BFO, BPNN in combination with MFCC. But the current research work does not support a live classification as the database is limited.

6. FUTURE SCOPE

The future aspects of this research work might involve the training with more number of datasets to check the performance of the system with other words also. The more training speech data used in a system, the better and higher the system's performance could be obtained. The accuracy of the identification process can be influenced by certain factors such as different level of surrounding noise during the recording session, the quality of the microphone used to input the speech signals and many others factors. Steps should be taken to minimize all these factors. Hybridizing ANN with BFOA can also be considered as future work.

7. REFERENCES

[1] M. Sher, N. Ahmad, M. Sher, "TESPAR Feature Based Isolated Word Speaker Recognition System" Proceeding of the 18th International Conference on Automation & Computing, 2012.

[2] W. Gevaert, G Tsenov, V Mladenov, "Neural Networks used for Speech Recognition" Journal of Automatic Control, Vol. 20, Issue. 7, 2010.

[3] A. Hossain, M. Rahman, U Prodhan, F. khan, "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition", International Journal of Information sciences and techniques(IJIST), Vol. 3, Issue.4, 2013.

[4] N. Srivastava, H. dev, S. Abbas, "Speech Recognition using MFCC and Neural Networks", International Journal of Engineering Science and Technology (IJEST), 2012.

[5] C. Ittichaichareon. S. Suksri, T. Yingthawornsuk, "Speech recognition using MFCC", International Conference on Computer Graphics, Simulation and Modeling (ICGSM),2012.

[6] P. Dalmiya, S.Dharun, P. Rajesh, "An Efficient Method for Tamil speech Recognition using MFCC and DTW for Mobile Applications", Proceedings of IEEE Conference

on Information and Communication Technologies (ICT), 2013.

[7] C. Sajjan, C. Vijaya, "Comparison of DTW and HMM for Isolated Word Recognition", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (IEEE), 2012.

[8] L. Muda, M. Begam, I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol. 2, Issue. 3, 2010.

[9] A. Seddik, M. Adawy, A. Shahin, "A Computer- Aided Speech Disorders Correction System for Arabic Language", 2nd International Conference on Advance in Biomedical Engineering (IEEE)", 2013.

[10] C. Yang, W. Yang, S. Wang, "Based on Artificial Neural Networks for Voice Recognition Word Segment", IEEE, 2011.

[11] K. Darabkh, A. Khalifeh, B. Bathech, S. Sabah, "Efficient DTW- Based Speech Recognition System for Isolated Words of Arabic Language", World Academy of Science, Engineering and Technology. Vol. 7, 2013.

[12] D. Paul, R. Parekh, "Automated Speech Recognition Of Isolated Words using Neural Networks", International Journal of Engineering Science and Technology (IJEST), Vol. 3, Issue. 6, 2011.

[13] B. Das, R. Parekh, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with neural Network Classifier", International Journal of Modern Engineering research (IJMER), Vol. 2, Issue. 3, pp-854-858, 2012.

[14] I. Al-Hadi, S. Hashim, "Bacterial Foraging Optimization Algorithm For Neural Network Learning Enhancement", International Journal of Innovative Computing, Vol. 1, Issue. 1, 2014.