INCREASING THE PERFORMANCE OF SPEECH RECOGNITION SYSTEM BY USING DIFFERENT OPTIMIZATION TECHNIQUES TO REDESIGN ARTIFICIAL NEURAL NETWORK


[2]Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, India
[3]Department of Electrical Engineering, Central University of Haryana, Mahendergarh, India

[1]shilpi.avighna@gmail.com, [2]ermadhu2003@gmail.com, [3]rajeshk_dubey@yahoo.com

ABSTRACT

Speech recognition has a high complexity and a broad range of applications, since it has to predict the word under many types of distortions. This paper aims to compare the performance of different optimization techniques like genetic algorithm, particle swarm optimization and artificial bee colony for optimizing the different hidden layers and neurons of the hidden layers of artificial neural network, for maximum speech recognition accuracy. The features of the input speech signals are extracted using amplitude modulation spectrogram. The outcome demonstrates that the accuracy of artificial bee colony is 95.3% and it performs better when compared with the other optimization techniques.

Keyword: Amplitude Modulation Spectrogram, Artificial Neural Network, Artificial Bee Colony, Speech Signal, Particle Swarm Optimization

1. INTRODUCTION

Speech processing has become very popular to interface with electronic gadgets [1-2]. In speech signal processing, the input speech signal is converted into its specific form called features that can be used by learning algorithms for the recognition of the speech signal. Speech quality is a major contributor to the end user’s perception of quality of service. As networks become more heterogeneous, complex and new technologies interoperate with legacy equipment, identifying the root cause of voice quality problems can be a challenging task [3-8]. For telephone service providers the evaluation and assurance of speech quality is a critical parameter.

In this work, the input speech signal is pre-emphasized by pre-emphasis filters then processed through first-order filters for smoothing the signal [9]. The detection of speech emotion is particularly useful for applications which require usual man–machine trades like web films, and computer tutorial applications [10-11]. The speaker recognizable proof framework can be executed by watching the voiced/unvoiced segments or by analyzing the speech energy appropriation. Such frameworks can be partitioned into two fundamental advances: feature extraction and speaker classification [12]. A model adaptation technique leaves the perceptions unchanged and rather refreshes the model parameters of the recognizer to be more illustrative of the speech [13-21]. The amplitude modulation spectrogram (AMS) design has been considered as a novel approach for classification. The AMS patterns can provide a dependable separation between voiced speech and noise as well as...
between unvoiced speech and noise [22]. A mainstream approach in speech innovation is artificial intelligence (AI) [23]. The classification of speech signal is imperative phenomenon in speech recognition process. In [24] Artificial neural network (ANN) system is utilized for classification. The ANN has been characterized from multiple points of view by a few researchers. The neurons of ANN are trained using input–output data sets [25], where, the isolated active words of speech are converted into text. For this, various words spoken by different persons are considered as the inputs. The features of input speech signals are extracted using AMS. The extracted input features are fed as the input to the proposed ANN for training and testing. Various optimization techniques have used to increase the efficiency of designs by choosing best solution among various viable solutions in different domains [26-41]. In [34] hybrid algorithms are used in breaking the hard electronic cipher text for linear and non-linear system. For optimizing hidden layer and neurons, various optimization techniques are utilized such as genetic algorithm (GA) [35], particle swarm optimization (PSO) [36-40] and artificial bee colony (ABC) [41].

This work demonstrates the use of ABC optimization algorithm for the purpose of redesigning the ANN structure and using it for increasing the speech recognition accuracy. Here, the ANN structure is redesigned by using existing Levenberg-Marquardt (LM) training algorithm [42] to retrieve optimal prediction rate accuracy. In the last, the result of ABC optimization method is also compared with other optimization algorithms like GA and PSO.

The organization of this paper is as follows. The proposed methodology related to speech recognition is given in Section 2. In Section 3 brief description of ANN is given. Section 4 deals with the brief description of optimization algorithms (ABC, PSO and GA) used in this paper. Section 5 describes the results, while, Section 6 concludes the paper.

2. PROPOSED METHODOLOGY

In this work the input isolated speech signal is recognized using ANN. Different words spoken by different persons (in ambient conditions) represent the normal speech signal and they are taken as the input speech signal. The features of input speech signals are then extracted by using AMS and these features are used as input to ANN for training and testing. ANN is utilized here for predicting the isolated word. For training and testing LM algorithm is used. The optimization techniques GA, PSO and ABC are used to optimize the neurons and the number of hidden layers. The complete block diagram of proposed work is shown in figure 1. This work is implemented in the working platform matrix laboratory (MATLAB 2015a).

The distinctive words are recorded by the different people (70 males, 50 females and 25 children) in ambient conditions. These words are used as input speech signal. In the current work AMS technique is used for extracting the features of the input speech signal. The AMS technique uses modulation spectrogram instead of spectral envelope as the information in the envelope is robust against noise and room reverberations. The three dimensional coefficients of AMS i.e. time, acoustic and modulation frequencies are obtained by firstly taking the short time fourier transform (STFT) of the input speech signal and then the envelope of the spectrogram is taken by squaring the magnitude of the complex values. The Bark scale decomposition is used for decaying the frequency scale into a set of critical bands and then second STFT is used for the analysis of the long-term spectral envelope of each sub-band.

3. ARTIFICIAL NEURAL NETWORKS

The artificial neural network includes three layers, for example, the input layer, hidden layer and the output layer in figure 2. The neurons of the various layers are connected to each other by synaptic weights. The synaptic weights are continuously updated during the learning or training process until an optimum solution is obtained.

The fundamental function of hidden neurons is given in equation (1).

\[ Q_j = \sum_{i=1}^{N} P_i \times w_{ij} \]  

where \( Q_j \) is a basics function, \( P_i \) is the input, \( w_{ij} \) is an input layer weight and \( i \) is a number of input. The activation function of hidden neurons is evaluated with equation (2).

\[ \tan{\text{sig}}(Q_j) = \frac{2}{(1 + \exp(-2Q_j)) - 1} \]
The LM algorithm is utilized to train and predict the improved artificial network structure design. It is a very popular optimization algorithm which uses both the gradient descent method and Gauss–Newton method for finding the optimum value.

The gradient vector is represented in equation (3).

\[
p = \begin{bmatrix}
  \frac{\partial S}{\partial w_1} \\
  \frac{\partial S}{\partial w_2} \\
  \vdots \\
  \frac{\partial S}{\partial w_n}
\end{bmatrix}
\]  (3)

Where \( S \) is the error of the network for that model and \( w \) suggests the weights. The Jacobian matrix obtained is given in equation (4).

\[
JC = \begin{bmatrix}
  \frac{\partial S_1}{\partial w_1} & \frac{\partial S_1}{\partial w_2} & \cdots & \frac{\partial S_1}{\partial w_n} \\
  \frac{\partial S_2}{\partial w_1} & \frac{\partial S_2}{\partial w_2} & \cdots & \frac{\partial S_2}{\partial w_n} \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{\partial S_m}{\partial w_1} & \frac{\partial S_m}{\partial w_2} & \cdots & \frac{\partial S_m}{\partial w_n}
\end{bmatrix}
\]  (4)

In this work for ANN structural design, different optimization techniques: GA, PSO and ABC are utilized.

4 OPTIMIZATION ALGORITHMS

4.1 Artificial Bee Colony Optimization

This optimization algorithm was proposed by Karaboga [43] and it is based on the intelligent foraging behaviours of swarms of honey bees. It consists of employed bees and unemployed bees. An unemployed bee comprises of onlookers bees (OB) and scout bees (SB). Based on each solution the neural network structure is designed and in this structure, features are given as input and then the output is predicted. Employed bees (EB) are used to find the new solution in defined solution \( i \), by the following equation (5)

\[
t_j = u_j + \Phi_j(u_j - u_k)
\]  (5)

Where, \( j \) is chosen randomly between \([1, C]\) where \( C \) is the no. of dimensions, \( \phi_j \) is distributed randomly between \([-1, 1]\) and \( ij \) demonstrates the record position and \( k \in \{1, 2...\text{size (i)}\} \) is also decided randomly. The critical limitation in this procedure is that \( k \neq i \). The probability \( (pi) \) is calculated for the \( uij \) according to the equation (6).

\[
p_i = \frac{F_i}{\sum_{i=1}^{N} F_i}
\]  (6)

Where, Fi is the fitness function. The above equation (6) finds the probability rate of fitness function after the greedy selection process. Based on the probability rate estimation, the solution is arranged and viewed as the initial solution for the onlooker bee section. Figure 3 gives the pseudo code of ABC algorithm.

4.2 Particle Swarm Optimization

In PSO, the single solution in the search space is called a particle. The fitness function given by equation (6) is optimized and the pbest (best solution) and gbest (global best of any particle in the population) values are updated in every iteration.

The velocity and positions of the particle are updated by the following equations (7-8)

\[
x[i+1]=x[i]+a_1*r_1*(pbest[i]-p[i])+a_2*r_2*(gbest[i]-p[i])
\]  (7)

\[
p[i+1]=p[i]+x[i]
\]  (8)

where \( x[i] \) is the particle velocity \( p[i] \) is the current particle (solution)
\( r1 \) and \( r2 \) are random numbers between \((0,1)\) and \( a1 \) & \( a2 \) are the learning factors.

4.3 Genetic Algorithm Optimization

GA [26, 35 and 36] is a global search examination technique for finding out the true or
approximate solution for an optimization problems.

The steps involved in GA:
1) A random initial population is generated.
2) GA then creates a new population by using individuals in the current generation to generate next population.
3) For generating new population
   a) Fitness value is computed by scoring each member of the current population. This is called raw fitness score.
   b) Expected values are then calculated by scaling the raw fitness scores to convert it into more usable range of values.
   c) Members called parents are selected based on their expectation.
   d) Elite individuals are selected from the individuals having lower fitness values in the current population 
   e) Next children are produced from parents either by making random changes to a single parent that is mutation or by crossover that is combination of the vector entries of a pair of parents.
   f) Current population is then replaced with the children from next generation.
4) GA then stops after meeting the stopping criteria.

5. RESULTS AND DISCUSSION

In the speech recognition process different speech signals from diverse persons are considered shown in Table 1 and the ANN classifying technique is used to predict the text. For training purpose, LM algorithm is performed and for optimizing hidden layer and neurons three optimization algorithms are applied named as PSO, GA and ABC.

Figure 4 shows network structures based on three algorithms, hidden layers and neurons are optimized for better recognition accuracy. For ABC algorithm 375 input layers, 2 hidden layers and 1 output layer are considered. In hidden layer 1 the neurons are 23 and in hidden layer 2 the neurons are 21.

For GA in hidden layer 1 there are 7 neurons and in hidden layer 2 the neurons are 20. For PSO the input and output layer is general and in hidden layer 1 the neurons are 8 and in hidden layer 2 the neurons are 30.

The comparison of the algorithms is done on the basis of sensitivity, specificity and accuracy given by equation (9-11)

\[
\text{Sensitivity} = \frac{TP}{TN + FP} \\
\text{Specificity} = \frac{TN}{TN + FP} \\
\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN}
\]

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

Table 2 gives the comparison based on classification time consumption for these three algorithms. The comparison graph is plotted with performance matrix such as accuracy, sensitivity and specificity in Table 3.

It is concluded from Table 2-3 that the ABC algorithm used here has maximum accuracy, specificity and sensitivity and takes minimum classification time as compared to GA and PSO algorithms.

6. CONCLUSION

In this work the recognition of the input speech signal is done using AMS as the feature extraction technique. The features extracted are then the input to ANN where LM algorithm is used for training. The existing structure of ANN is optimized using ABC, PSO and GA. The outcome clearly demonstrates that the proposed method of speech recognition is superior to different existing methods. This method decreases the complexity of the calculation and offers good recognition result along with reduction in time consumption. The accuracy value of Artificial Bee Colony is 95.3%, while contrasted with other algorithms like PSO and GA. In future, the strategy can be connected to other speech recognition systems for increasing the efficiency.

REFERENCES

[1] C.Vimala, V.Radha, "A Review on Speech Recognition Challenges and Approaches", in World of Computer Science and
Information Technology Journal (WCSIT),


M. Jain, M. Gupta, M. Alabbas, H. Abdulkareem, A. kareem,


A. Kumar, D. Kumar, S. K. Jarial, “A Review on Artificial Bee Colony
Algorithms and Their Applications to Data Clustering", Cybernetics and Information Technologies, vol 17, no 3,2017


Fig. 1 Overall block diagram

Input Speech Signal

Feature Extraction using Amplitude Modulation Spectrogram (AMS)

Artificial Neural Network (ANN) training using Levenberg-Marquardt (LM) training algorithm

Optimization of the hidden layers and neurons using Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA)

Fig. 2 Neural network structure

Input Layer

Hidden Layer (neurons with a random weight $w_{11}$, $w_{12}$, ..., $w_{ij}$)

Output Layer
Notations used:
FS - Food Sources
NS – New Solution
iter – iteration
max_iter – maximum iterations
R – Random number
bsn – best solution

<table>
<thead>
<tr>
<th>Inputs: Max Cycles, Colony Size, limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Initialize FS</td>
</tr>
<tr>
<td>2: Evaluate the initialized FS</td>
</tr>
<tr>
<td>3: iter =1</td>
</tr>
<tr>
<td>4: Set the initial FS to the best Sn solutions</td>
</tr>
<tr>
<td>5: while iter &lt; max_iter do</td>
</tr>
<tr>
<td>6: Generate NS using EB</td>
</tr>
<tr>
<td>7: Evaluate NS &amp; apply greedy selection process</td>
</tr>
<tr>
<td>8: Calculate probability using eq. (6)</td>
</tr>
<tr>
<td>9: Generate the NS using OB</td>
</tr>
<tr>
<td>10: Evaluate NS &amp; apply greedy selection process</td>
</tr>
<tr>
<td>11: NS generated and abandoned solutions determined randomly using SB</td>
</tr>
<tr>
<td>12: Save best solution</td>
</tr>
<tr>
<td>13: iter=iter+1</td>
</tr>
<tr>
<td>14: End while</td>
</tr>
<tr>
<td>15: Return bsn</td>
</tr>
<tr>
<td>Output: bsn</td>
</tr>
</tbody>
</table>

Fig. 3 Pseudo code of ABC algorithm
Fig. 4 Neural network structure for ABC, GA and PSO algorithms

a) ABC Neural Network Structure

b) GA Neural Network Structure

c) PSO Neural Network Structure

Table 1 Sample of different person’s input signal data
<table>
<thead>
<tr>
<th>Input signals</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lion</td>
<td><img src="image1" alt="Lion Person 1" /></td>
<td><img src="image2" alt="Lion Person 2" /></td>
<td><img src="image3" alt="Lion Person 3" /></td>
</tr>
<tr>
<td>Lotus</td>
<td><img src="image4" alt="Lotus Person 1" /></td>
<td><img src="image5" alt="Lotus Person 2" /></td>
<td><img src="image6" alt="Lotus Person 3" /></td>
</tr>
<tr>
<td>Peacock</td>
<td><img src="image7" alt="Peacock Person 1" /></td>
<td><img src="image8" alt="Peacock Person 2" /></td>
<td><img src="image9" alt="Peacock Person 3" /></td>
</tr>
<tr>
<td>Rose</td>
<td><img src="image10" alt="Rose Person 1" /></td>
<td><img src="image11" alt="Rose Person 2" /></td>
<td><img src="image12" alt="Rose Person 3" /></td>
</tr>
<tr>
<td>Sunflower</td>
<td><img src="image13" alt="Sunflower Person 1" /></td>
<td><img src="image14" alt="Sunflower Person 2" /></td>
<td><img src="image15" alt="Sunflower Person 3" /></td>
</tr>
</tbody>
</table>
Table 2 Comparison of Classification Time consumption

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time consumption (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>12.2 seconds</td>
</tr>
<tr>
<td>PSO</td>
<td>13.5 seconds</td>
</tr>
<tr>
<td>GA</td>
<td>15 seconds</td>
</tr>
</tbody>
</table>

Table 3. Comparison Results of different algorithms

<table>
<thead>
<tr>
<th>Optimizing Algorithm</th>
<th>ANN-ABC</th>
<th>ANN-PSO</th>
<th>ANN-GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>95.3</td>
<td>93.3</td>
<td>90</td>
</tr>
<tr>
<td>Specificity</td>
<td>97.7</td>
<td>96.7</td>
<td>94.3</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>88.3</td>
<td>87.7</td>
<td>87.2</td>
</tr>
</tbody>
</table>