International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Speech Recognition with Advanced Feature Extraction Methods Using Adaptive Particle Swarm Optimization

Bright Kanisha[1*] **Ganesan Balarishnanan**[2]

[1]*Indra Ganesan College of Engineering, Trichirappalli, India*
[2]*Indra Ganesan College of Engineering, Trichirappalli, India*
Corresponding author's Email: kanishab0880@gmail.com

Abstract: Nowadays, speech recognition applications are becoming increasingly effective. In the market, different interactive speech aware applications are obtainable. In this work, from the input speech signal by recognizing the content involves three stages such as the preprocessing, feature extraction and Multi Support Vector Machine (SVM). The signal is processed and noise free signal is produced by processing the signal and the features are extracted. For optimize these features different optimization algorithms are utilized. From this algorithm the optimal features such as peak signal frequency, Tri-spectral feature, and discrete wave transform (DWT) attain the APSO technique. These optimal features are given as the input of the multi SVM and the signal in testing process, the signal correctly recognize the text. From the results the optimization algorithm (APSO) obtains the 97.8% accuracy compared to the existing technique SVM linear kernel function.

Keywords: Speech recognition; Mel Frequency Cepstral Coefficients; Tri spectral feature and discrete wave transform; Kernel function; feed forward back propagation; Multi-class SVM; adaptive particle swarm optimization.

1. Introduction

Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal [1]. Signal processing transforms the input speech signal into a form that can be processed by recognizer [2]. Speaker identification system can be implemented by observing the voiced/unvoiced components or by analyzing the speech energy distribution [3]. Model adaptation methods leave the observations un-changed and instead updates the model parameters of the recognizer to be more representative of the observed speech [4]. Since speech recognition has to be performed in different environmental conditions, therefore, the features extracted must also be robust to background noise and sensor mismatch conditions [5]. Speech emotion recognition is particularly useful for applications which require natural man–machine interaction such as web movies and computer tutorial applications where the response of those systems to the user depends on the detected emotion [6]. Speech recognition is a convenient basis for the development of human-machine interfaces, telecommunication services, and multimedia tools, either as a stand-alone tool as the input (e.g., data entry) or for further natural language processing (e.g., spoken language translation) [7].

Applying convolution of speech features, which often happen between different speakers and for the same speaker in different methods [8]. Over the last few years, speech synthesis research has moved from using unit selection speech synthesis technology [9]. Digital speech is different from audio signal in respect to factors like production model, perception, bandwidth, loudness, and intensity. Digital watermarking is the proper technique to protect and monitor the digital media [10]. In the recent years, the emotion recognition from speech has noticeable applications in the speech-processing systems, such as spoken tutoring systems, medical emergency domain to detect stress and pain, interactions with robots, computer games, and call centers [11]. The main objective of this research work investigate the speech signals classification in SVM process with optimal features selected in feature extraction process. This optimal

feature collection process different encouraged swarm intelligence optimizations are used improve accuracy of classification process. In section 2 there is an elaborate description regarding the literary reviews. Section 3 is rich with colorful data on the proposed technique. In section 4 discuss the speech signals classification performance parameters compared with proposed and existing technique. Section ends with a befitting conclusion.

2. Literature Review

In 2010 A. Dev et al [12] deeply debated on the issue of augmenting the strength of speech front-ends and launched an innovative set of MFCC vector evaluated by means of three phases. In the first stage, the relative higher order autocorrelation coefficients were efficiently mined. Thereafter the magnitude spectrum of the consequent speech signal was assessed by means of the fast Fourier transform (FFT) and it was distinguished in terms of frequency. In the final phase, the distinguished magnitude spectrum was metamorphosed into MFCC-like coefficients, termed as MFCCs mined from the Differentiated Relative Higher Order Autocorrelation Sequence Spectrum (DRHOASS).

In 2011 F. L. Huang [13] have proposed an effective approach for Chinese speech recognition on small vocabulary size was independent speech recognition of Chinese words based on Hidden Markov Model (HMM). The features of speech words are generated by sub-syllable of Chinese characters. Total 640 speech samples are recorded by 4 native males and 4 females with frequently speaking ability. The preliminary results of inside and outside testing achieve 89.6% and 77.5%, respectively. The final precision rates for inside and outside test in average achieve 92.7% and 83.8%. The results prove that the approaches for Chinese speech recognition on small vocabulary are effective.

In 2013 Poonkuzhali et al [14] have proposed the Speech was one of the most promising models by which people can express their emotions like anger, sadness, and happiness. Acoustic parameters of a speech signal like energy, pitch, Mel Frequency Cepstral Coefficient (MFCC) was important in find out the state of a person. The features get reduced to 16.6% in 300 iterations. Ant Colony Optimization is able to select the more informative features without losing the performance.

In 2014, I. E. Henawy et al [15] have proposed the goal of speech recognition produce a machine which will recognize accurately the normal human speech from any speaker. A recognition rate of 98% was obtained using the proposed feature extraction technique. The features based on the Cepstrum give accuracy of 94% for speech recognition while the features based on the short time energy in time domain give accuracy of 92%. The features based on formant frequencies give accuracy of 95.5%. It is clear that the features based on MFCCs with accuracy of 98% give the best accuracy rate. So the features depend on MFCCs with HMMs may be recommended for recognition of the spoken Arabic digits.

3. Proposed Methodology

The proposed speech recognizing methodology in analysis of the speech signal by predicting the content involves three stages such as the preprocessing, feature extraction and Multi support vector machine (SVM).

In this paper incorporates the existing work features such as input speech signal contrasted with the standard speech signal, Peak frequency modulation, MFCC, Tri spectral features and Discrete Wavelet transform (DWT). DWT process is applied on the wavelet transform in the input speech signal some measures are obtained. For optimizing the above mention features optimization techniques are used in speech recognition process. The optimization algorithm such as genetic algorithm (GA), adaptive genetic algorithm (AGA), particle swarm optimization (PSO), Harmony Search (HS) and adaptive particle swarm optimization (APSO) are utilized to obtain the optimal features. In this optimal features are given to the input of the multi SVM process and the SVM process linear kernel function are utilized to predict the text in speaker dependent process.

Figure 1 shows the block diagram of our proposed speech recognizing methodology.

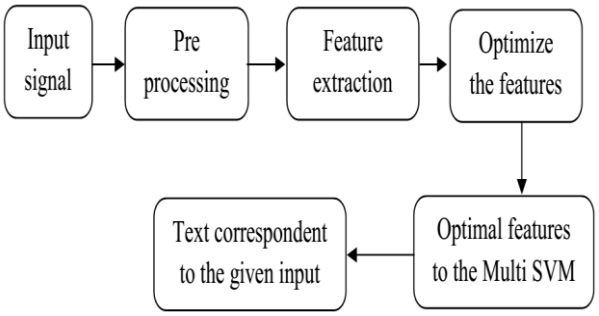

Figure.1 Block diagram for proposed work

3.1 Input Signal

The different voice is recorded by the different persons and all are spoken the same words like dove, rose etc. It is used as the input signal *S(i)*.

3.2 Pre Processing

The input signal is changed as vector and the variation found. Usually, for the noise removal process, a Gaussian filtering is employed, which is defined as *P(k)* where y represents the noise and σ denotes a deviation.

$$P(k) = \exp\left(\frac{-y^2}{2y\sigma^2}\right) \tag{1}$$

3.3 Feature Extractions

The feature extraction process involves the analysis of the speech signal. For speech signal feature extraction, the spectral analysis technique is used. Feature extraction means the transformation of the input data into the set of features. In this work consider the features as

- Compare the standard speech signal and normal speech signal
- Peak frequency modulation
- MFCC
- Tri spectral features
- Discrete Wavelet transform(DWT)

3.3.1. Compare the standard speech signal and normal speech signal *C(k)*

Normal speech signals are contrasted in feature extraction with the standard speech signal. A frequency range of threshold is then located for the specific speech signal.

$$C(k) = S(s) - P(s) \tag{2}$$

Where, *S(s)* is a standard signal, *P(s)* is a Normal signal

3.3.2. Peak frequency modulation *f(k)*

The feature extraction with peak signal extraction and next the input signal is contrasted with the standard signal which creates main impact in the output signals.

$$f(k) = \begin{cases} p+1 & A > T \\ 0 \end{cases} \tag{3}$$

Where, p is a peak value, A is amplitude and T is a threshold value.

3.3.3. Mel frequency cepstral coefficients (MFCC)

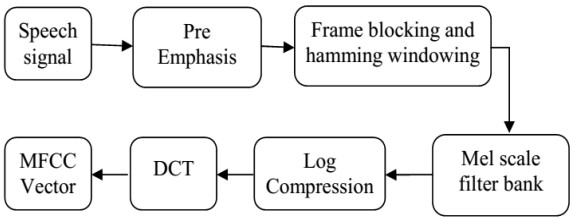

Figure.2 MFCC process

Mel frequency cepstral coefficients (MFCC) is one of the most victorious feature representations, commonly employed in automatic speech and speaker recognition in speech recognition associated tasks and using a filter bank analysis the coefficients are attained.

Pre-emphasis, frame blocking, windowing, filter bank analysis, logarithmic compression, and discrete cosine transformation are the steps occupied in the feature extraction.

The overall process of the MFCC feature computation is illustrated in Figure 2.

Pre-Emphasis

Pre-Emphasis is the development to increase the magnitude of frequencies regarding the magnitude of other frequencies. At first, the speech signal is pre-emphasized by a first order FIR filter with pre-emphasis coefficient β. The first order FIR filter transfer function in the z domain is,

$$E(z) = 1 - \beta z^{-1} \tag{4}$$

The pre-emphasis coefficient β lies within the range $0 \le \delta \le 1$.

$$e(v_i^r{}') = \rho(v_i^r{}') - \beta\rho(v_i^r{}'-1) \tag{5}$$

$$e(v_j^t{}') = \rho(v_j^t{}') - \beta\rho(v_j^t{}'-1) \tag{6}$$

For training the word equation (5) is used and for testing the word equation (6) is used. Where pre-emphasis coefficient is represented by β, *v* in equation (5) represents training word and *v* in equation (6) represents testing the word.

Frame blocking and hamming windowing

The arithmetic features of a speech signal are invariant just with in short time intervals. Now, the adjacent frames being separated by f_A samples (frame shift) and the pre-emphasized signal is blocked into frames of f_s samples (frame size). If the k^{th} frame of speech is $x_k(v_i^r{}')$, $x_k(v_j^t{}')$ and there are *K* frames within the entire speech signal, then

$$x_k(v_i^r{}') - \rho(f_{A^r} + v_i^r{}'), \quad 0 \le v_i^r{}' \le f_{A^r} - 1 \tag{7}$$

In windowing, each of the above frames is multiplied with a hamming window so as to keep

stability of the signal to decrease the signal discontinuities at the start and end of the frames while each frame is windowed. Windows are selected to tape the signal at the edges of each frame. The hamming window equation is specified as,

$$H(n) = x(n) \times r(n) \tag{8}$$

Where, *r(n)* is a window function

$$r(n) = 0.54 - 0.46\cos\left(2\prod\frac{n}{N}\right)$$
$$0 \le n \le N \tag{9}$$

Mel scale filter bank

The filter bank analysis is carried out to change each time domain frame of f_s samples into frequency domain. The filters are collectively called as a Mel scale filter bank and the perceptual processing prepared within the ear is inspired by the frequency response of the filter bank.

$$F(Mel) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{10}$$

Discrete Cosine Transformation (DCT)

This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT).The set of coefficient is called acoustic vectors. Then, Discrete Cosine Transform (DCT) is applied to the filter outputs and of a specific speech frame the first few coefficients are grouped together as a feature vector. The K^{th} MFCC coefficient can be given as,

$$D(k) = c(k)\sum_{n=1} x(n)\cos\left(\frac{\prod}{2N}(2n-1)(k-1)\right)$$
$$k = 1,2,..N \tag{11}$$

Where

$$c(k) = \begin{cases} \frac{1}{\sqrt{N}} & k = 1 \\ \sqrt{\frac{2}{N}} & 2 \le k \le N \end{cases} \tag{12}$$

3.3.4. Tri spectral features

To apply tri-spectral analysis for the classification of the speech signal is the purpose of this section. Normally the speech signals are recorded and next the tri spectrum features are examined. Tri-spectrum is a kind of statistics employed to recognize the output speech signal which is not directly relative to the consequent input speech signal. Steps occupied in tri spectrum:

Step1

Input signal based take the amplitude values

Step 2

Maximum amplitude apply the fast Fourier transform

$$nfft = \max(nfft, 2^A) \tag{13}$$

Step3

Calculate the frequency response for the input signal X_f and calculate the complex conjugate for the frequency response X_{fc}.

Step 4

X_f and X_{fc} based find the tri spectrum output

$$T_{spec} = X_f \times (X_f \times X_f^T) \times hankel(X_{fc}) \tag{14}$$

$$T_{spec} = FFTshfit(T_{spec}) \tag{15}$$

Step 5

Tri spectrum output based finds the features

- By using the number of pixel present in the row, column, and both diagonals we can calculate the centre point of the pixel.
- The properties of the region near the centre point of tri-spectrum can be calculated by using the region props function in the Matlab. The region props function helps to evaluate some properties of the region.
- To estimate the tri-spectrum of a signal the following processing steps are performed

Location of Maximum Values

The location of maximum values can be obtained by computing the maximum values from both row and column. Thus we get two maximum values from both horizontal h_{max} and vertical line v_{max}

Sum of Diagonal Values

In this calculate the sum of all the values present in the diagonals from both the direction i.e. left diagonal DL_{sum} and right diagonal DR_{sum}.

Sum of Centre Column CC_{sum} and Centre Row RC_{sum} Values

This value can be evaluated by computing the sum of the values present in the centre row RC_{sum} and the values present in the centre column. CC_{sum}

Orientation of Region

The orientation of the image O_r is obtained by measuring the angle between the horizontal axis and the major axis of the ellipsoid.

Eccentricity of Region

The eccentricity of the region T_r can be calculated by taking the relation of distance between the foci of ellipse and the major axis.

Solidity of Region

Solidity is the ratio of region area to the convex area of the region. The solidity of the region is calculated using the formula

$$S_r = \frac{A_r}{CA_r} \tag{16}$$

Where, A_r is the Area of the region, CA_r is the convex area of the region, S_r is the solidity.

Extent of Region

Extent of the region E_r is measured by taking the relation of region area to the bounding box area which is given by the formula

$$E_r = \frac{A_r}{BA_r} \tag{17}$$

Where, A_r is the Area of the region, BA_r is the convex area of the region, E_r is the extent.

Perimeter of Region

The perimeter of the region P_r can be measured by taking the number of neighboring pixel of the region and calculating the space between the adjacent pixels which lies in the border of the region.

Entropy of Tri Spectrum

Entropy is determined by taking the probability of a process or information content. The entropy for trispectrum E_s is calculated using the formula

$$E_S = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{18}$$

Where, $P(X_i)$ is the probability mass function

3.3.5. Discrete wavelet transforms (DWT)

Wavelet transform offers a framework in which a signal is decayed, with each level related to a coarser resolution or lower frequency band, and higher frequency bands. Two most important groups of transforms are there, continuous and discrete. The DWT, which applies a two-channel filter bank (with down sampling) iteratively to the low-pass band (initially the original signal).

For the speech signal the wavelet transform is given as

$$D(p,q) = \int_{-\infty}^{\infty} s(t).\psi p, q(t) dt \tag{19}$$

Where, $\psi p, q(t)$ is the wavelet function.

Two dimensional Haar wavelet transform because it reduces the computational time and also it extracts more features. For the t input speech signal φ_t the haar wavelet transform P_t is given as

$$p_t = H_t \phi_t \tag{20}$$

Then the mean value of the coarse coefficients is calculated by taking the average of the coarse coefficient.

$$C[m_t] = \omega_{m_t} \tag{21}$$

Where ω_{mt} the mean value for approximation coefficient

From the mean value the standard deviation of the coarse coefficients is measured by taking the square root of the mean value.

$$\sigma_{a_t} = \sqrt{C[m_t] - (C[m_t])^2} \tag{22}$$

Where, σ_{mt} is the standard deviation for approximation coefficient

Speech recognition process optimizes the above mention features and the dissimilar optimization algorithms are employed. The optimal features such as peak frequency modulation, DWT and tri spectral features are attained in the APSO algorithm.

3.4 Optimal feature selection using APSO algorithm

Based on the simulation of the social behavior of bird flocks a particle swarm optimizer is a population based stochastic optimization algorithm modeled. PSO is a population-based search process where individuals initialized with a population of random solutions, referred to as particles, are clustered into a swarm.

1. Initialize the solution (F_{ei})

$$F_{ei} = \{F_{e1}, F_{e2}, ...F_{en}\}$$

2. Find the fitness value (F_i)

$$F_i = \max(A)$$

3. Initialize the P_{best} *and* G_{best} value

4. Compute the acceleration factor $nc1$ *and* $nc2$

$$nc_1 = \frac{2}{3}(c_{1\max} - c_{1\min})\left(\frac{f_{\min}}{f_{avg}} + \frac{f_{\min}}{2f_{\max}}\right) + c_{1\min}$$

$$nc_2 = \frac{2}{3}(c_{2\max} - c_{2\min})\left(\frac{f_{\min}}{f_{avg}} + \frac{f_{\min}}{2f_{\max}}\right) + c_{2\min}$$

5. Calculate the velocity and update the position

$$V_i^d = w^d V_i^d + nc_1 r_1 (bp_i^d - p_i^d) + nc_2 r_2 (gp^d - p_i^d)$$

6. Find the fitness for updating solution

$$if(F_{enew}) > f(F_e)$$

7. Store the best solution so far attained

Iteration=Iteration+1

8. Stop until optimal solution attained

Pseudo code for APSO

3.4.1. Initialization

Initialization process considers the features as the input such as compare the standard speech signal and normal speech signal (F_{e1}), Peak frequency modulation (F_{e2}), *MFCC* (F_{e3}) Tri spectral features (F_{e4}) and Discrete Wavelet transform (F_{e5}). These feature based generate the particles randomly.

3.4.2. Fitness Function (F_i)

The fitness function chooses which should be used for the constraints according to the current population. In step 2 (F_i) mentions the maximum accuracy *(A)* for each word.

3.4.3. Initialize *gp* and *bp*

Initially the fitness value calculated for each particle is set as the P_{best} value of each particle. Among the P_{best} values, the best one is selected as the G_{best} value

3.4.4. Compute the acceleration factors

The acceleration factors are computed using the equation nc_1 and nc_2. In this equation c_{1max}, c_{1min} is a minimum and maximum values of c_1, f_{min}, f_{avg}, f_{max}- minimum, average and maximum fitness value of the particles and c_{2max}, c_{2min} is minimum and maximum values of c_2.

3.4.5. Velocity Computation and update the position

The new velocity is calculated using the below equation. Substitute the nc_1 and nc_2 values in the velocity equation V_i^d. Then calculate the fitness function again and update the P_b and G_b values. If the new value is better than the previous one, replace the old by the current one.

3.4.6. Optimal solution

Based on above mention process attain the optimal features the process will be continued. Thus the set of optimal features (F_{e2}, F_{e4}, F_{e5}). These features are obtained from the adaptive particle swarm optimization technique is given to multi SVM for training process and produce the high accuracy.

3.5 Multi support vector machine

SVM are fundamentally two-class classifiers which pursue the traditional way to do multiclass classification. The two-class linear classifiers can be made bigger to I >2 classes. The optimal features are specified to the input and linear kernel function is employed to categorize the text in SVM process.

$$(F_{e1(optimal)}, F_{e2(optimal)}, ...F_{en(optimal)}) \quad (23)$$

$$F_n \in R^D$$

However, to solve multiclass problems these are not well-designed approaches. By the construction of multiclass SVMs, an improved alternative is provided, where we construct a two-class classifier over a feature vector $\psi(\vec{F}_e, y)$ derivative of the pair consisting of the input features and the class of the datum. The classifier chooses the class test time the feature vector in equation (23) bias value based obtain the maximum value

$$y = agr\max_y, \vec{w}^T \psi(F_{e(optimal)}, y') \quad (24)$$

Automatically, a superior segregation is achieved by the hyper plane containing the highest distance to the nearby training data point of any class. Linear kernel equation (24) in multi SVM process build the hyper plane equation is successfully arranged. In numerous types of linear classifiers this universal technique can be employed to hand over a multiclass formulation.

$$W = \sum_{i=1}^{n} \alpha_i y_i F_{ei(optimal)} \quad (25)$$

$$C => WF_e + y \quad (26)$$

At last, the hyper plane based categorizes the signal by means of the (C) equation, in which the bias value and input (F_e) and hyper plane value are

effectively utilized to locate the class and forecast the related text.

4. Results and Discussion

The proposed speech recognition process with optimization technique for recognize the speech signal is done by using MATLAB software version 2014a latest version in a system having 8 GB RAM with 64 bit operating system having i5 Processor. In this utility based attains the parameters like the True positive (*TP*), True negative (*TN*), false positive (*FP*) and false negative (*FN*).

4.1 Statistical measures of the performance of different texts

To categorize the text, different texts are regarded for optimizing feature to be employed in the SVM process. Below table shows that the sample input text with performance analyzing process.

Table 1 shows that for input speech signal "dove" evaluates the TP, TN, FP, FN and from these attains the result such as the Sensitivity (Se), Specificity (Sp), Accuracy (A), False Positive Rate (FPR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Discovery Rate (FDR) and Mathews Correlation Coefficient (MCC) values as per the below-mentioned formulas, above table shows statistical parameter in APSO algorithm with SVM linear kernel function. In this word expect person 1 other persons will get the 100% accuracy and sensitivity, specificity and the text the TP has 1 TN has 9 all the statistical parameters are very best in person 2,3,4 and 5.

Formulas used for finding the vales in table 1.

$$Sensitivity = \frac{No.of\ TP}{No.of\ TP + No.of\ FN} \quad (27)$$

$$Accuracy = \frac{(TP\ +\ TN)}{TP\ +\ TN\ +\ FP\ +\ FN} * 100 \quad (28)$$

$$FPR = \frac{FP}{N} = \frac{FP}{(FP + TN)} \quad (29)$$

$$PPV = \frac{TP}{(TP + FP)} \quad (30)$$

$$NPV = \frac{TN}{(TN + FN)} \quad (31)$$

$$FDR = \frac{FP}{(FP + TP)} \quad (32)$$

$$MCC = \frac{(TPTN\text{-}FPFN)}{\sqrt{PNP'N'}} \quad (33)$$

Table 1. Statistical measures of Dove

Persons	Se	Sp	A	PPV	NPV	FPR	FDR	MCC
1	0	1	0.9	-	0.9	0	-	-
2	1	1	1	1	1	0	0	1
3	1	1	1	1	1	0	0	1
4	1	1	1	1	1	0	0	1
5	1	1	1	1	1	0	0	1

4.2 Comparison graphs for proposed method

Speech recognition process from the input signal some features are extorted, for optimize these features dissimilar optimization algorithm are employed. The algorithms such as GA, AGA, PSO, HS and APSO are used to attain the optimal features.

Figure 3 shows that the accuracy of the all the words in different optimization algorithms GA, PSO, AGA and APSO among these algorithms maximum accuracy attained in the APSO technique. In this optimization process APSO algorithm has the maximum accuracy as 97.8% and this is compared to the PSO and the accuracy has minimized as 3.4%. Then APSO compared with the AGA and the accuracy has minimized as 3.56% and the AGA the accuracy is 94% and finally compared to the GA the accuracy minimized as 5.8%. HS technique compared to proposed work the difference is 0.56%.The proposed APSO algorithms swarm behavior based optimizes the solutions to get the optimal features and comparison technique HS optimization technique inspired by the improvisation of Jazz musicians. Specifically, the process by which the musicians (who have never played together before) rapidly refine their individual improvisation through variation resulting in an aesthetic harmony. Totally the accuracy of the APSO compared to the other algorithm as 4.36% and the accuracy will be increased.

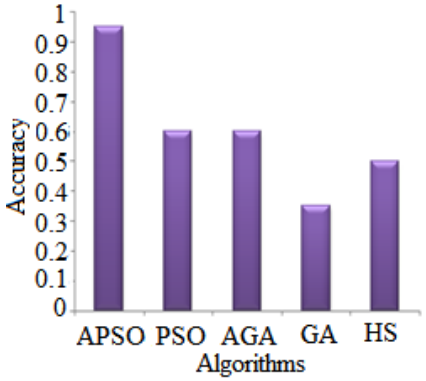

Figure.3 Comparison graph for different algorithm

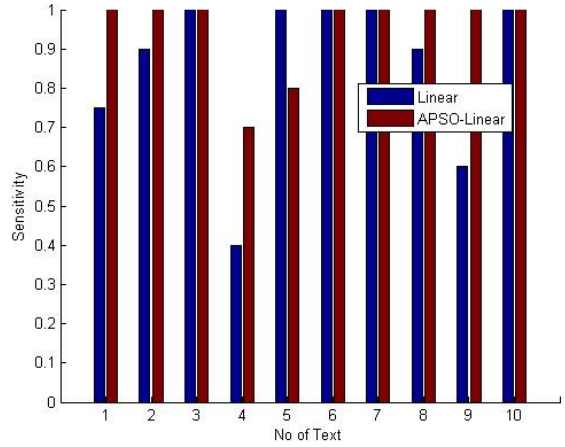

Figure.4(a) Comparsion graph for sensitivity

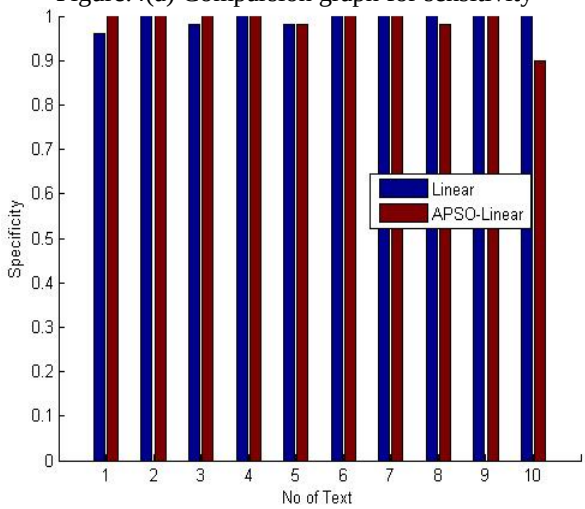

Figure.4(b) comparison graph for specificity

In figure 4 shows sensitivity and specificity of the different words such as Dove, eagle, jasmine, king fisher, lion, lotus, peacock, rose, sun flower and tiger. These words sensitivity and specificity is compared with the existing method SVM linear kernel function and the proposed optimization algorithm APSO with the SVM linear kernel function. Figure (a) shows the word dove the sensitivity of the linear kernel is 75% and is compared with the APSO the sensitivity is increased as 25%. Then the next word eagle also optimization algorithm obtained the 100% accuracy and is compared with the liner kernel function the sensitivity minimized as 92.3% and the net word both the method attain the 100% accuracy. Figure (b) shows the text signal the specificity of APSO linear kernel function is 100% and existing method linear kernel specificity is minimized as 3%. In the next text also the APSO linear kernel function when compared to the other technique is 2.25

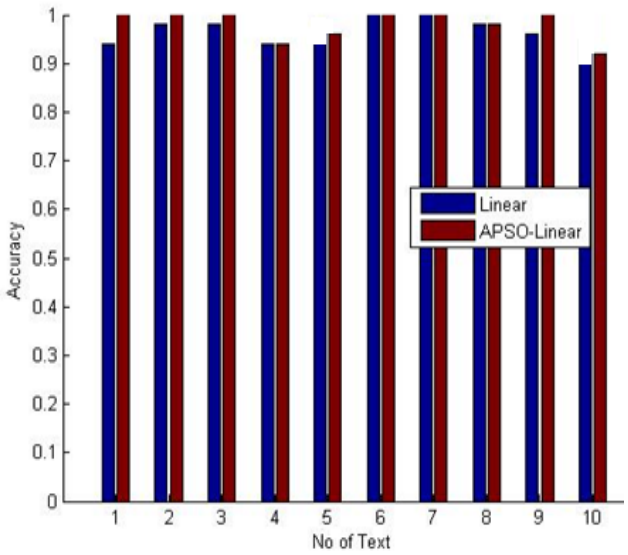

Figure.5 comparison graph for accuracy

Figure 5 shows the accuracy values for the different texts such as dove, eagle, jasmine, kingfisher, lion, lotus, peacock, rose, sunflower, and tiger. The comparison is performed for proposed and existing method linear function. For the text dove the accuracy in proposed is 100% and in linear function is 95% and the text signal eagle the accuracy in proposed is 99% and linear function is 97%. Similarly, other text signal also APSO with the linear function attain the 99% accuracy compared to the existing method.

Figure 6 shows the speech signal classification in proposed SVM and ANN techniques. The maximum accuracy of SVM compared to ANN technique the difference is 4.25%, similar difference in other performance measures like sensitivity and specificity.

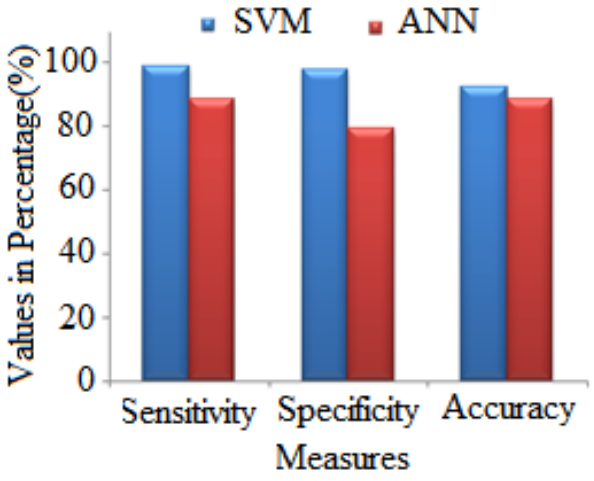

Figure.6 Comparative analysis for classification technique

4.3 Convergence Graph

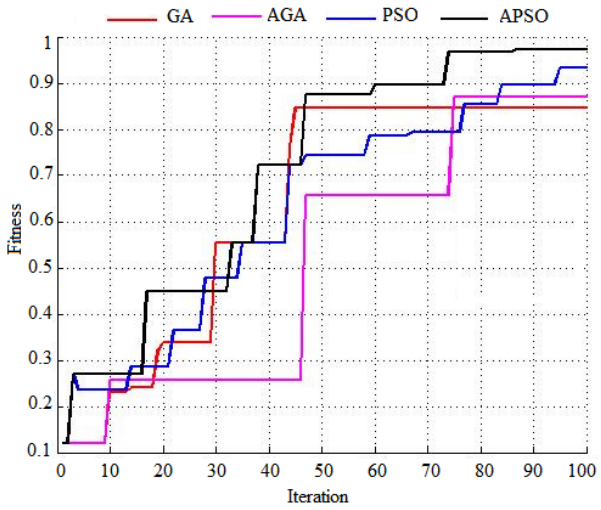

Figure.7 Convergence graph

Figure 7 demonstrates that the convergence graph is plotted between the iteration and fitness estimations of the different strategies, for APSO.AGA, GA, and PSO. This graph fundamentally resolves that the PSO procedure with the acceleration factor is specified the greatest fitness by means of least possible iteration. PSO Strategy gets the minimum iteration through the graph for presenting the ideal result. APSO the precision 98% reached in the 85 iteration, during the initial iteration the fitness value in APSO is 15 and other algorithm approximately in 11% and also the HS technique compared to APSO the difference is 2.3%. Through the graph the adaptive particle swarm Optimization strategy just states the ideal fitness value with the competent results.

5. Conclusion

In this document the performance of the speech signal based on the different features optimized employ to categorize the text where the Multi Support vector machine with the linear kernel function. The precision of the each person's is 97%, 99%, 98%, 98% and 97% in APSO technique. The above mentioned result clearly confirms that our suggested technique is better than the presented linear kernel function. The benefit of feature is that they decrease the difficulty of the calculation and present good recognition result along with diminution in time utilization. Therefore, in future, this recommended technique can be successfully applied to the other speech recognition process. The accuracy level has clearly evinced that the proposed algorithm is highly efficient in speech independent and dependent signal with best features.

Reference

[1] C. Ittichaichareon, S. Suksri and T. Yingthawornsuk, "Speech Recognition using MFCC", *International Conference on Computer Graphics*, pp. 135-138, 2012.

[2] K. Kumar, Aggarwal and A. Jain, "An Analysis of Speech Recognition Performance Based Upon Network Layers and Transfer Functions", *Journal of Computer Science, Engineering and Applications*, Vol. 1, No. 3, pp. 11-20, 2011.

[3] K. Daqrouqa and T. A. Tutunjiba, "Speaker identification using vowels features through a combinedmethod of formants, wavelets, and neural network classifiers", *Journal of Applied Soft Computing*, Vol. 27, pp. 231-239, 2015.

[4] M. Seltzer, D. Yu and Y. Wang, "An Investigation of Deep Neural Networks for Noise Robust Speech Recognition", *Journal of microsoft research*, pp. 7398-7402, 2013.

[5] A. Biswas, Sahu, A. Bhowmick and M. Chandra, "Articulation based admissible wavelet packet feature based on human cochlear frequency response for TIMIT speech recognition", *Journal of Ain shams Engineering*, Vol. 5, No. 4, pp. 1189-1198, 2014.

[6] M. ElAyadi, M. Kamel and F. Karray, "Survey on speech emotion recognition:Features, classification schemes, and databases", *Journal of Pattern Recognition*, Vol. 44, No. 3, pp. 572–587, 2011.

[7] G. Heigold, H. Ney, R. Schlüter and S. Wiesler, "Discriminative training for automatic speech recognition", *Journal of signal processing*, pp. 58-69, 2012.

[8] O. A. Hamid, L. Deng and D. Yu, "Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition", *Journal of computer science and engineering*, pp. 3366-3370, 2013.

[9] A. Black, T. Bunnell, Y. Dou, P. K. Muthukumar, F. Metze, D. Perry, T. Polzehl, K. Prahallad, S. Steidl and C. Vaughn, "Articulatory Features For Expressive Speech Synthesis", *Journal of audio speech and language processing*, pp. 4005-4008, 2012.

[10] M. A. Nematollahi, A. Haddad and F. Zarafshan, "Blind digital speech watermarking based on Eigen-value quantization in DWT", Journal of King Saud University Computer and Information Sciences, Vol.27, No. 1, pp.58-67, 2015.

[11] D. Gharavian, M. Sheikhan, A. Nazerieh and S. Garoucy, "Speech emotion recognition using FCBF feature selection method and GA-optimized fuzzy ARTMAP neural network", *Journal of neural computer and applic*, Vol. 28, No. 1, pp. 1-12, 2011.

[12] A. Dev and P. Bansal, "Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order

Autocorrelation Coefficients ", *Journal of Computer Applications*, Vol. 10, No. 8, pp. 36-38, 2010.

[13] F. L. Huang, "An Effective Approach for Chinese Speech Recognition on Small size of Vocabulary", *Journal of signal and image processing*, Vol. 2, No. 2, pp. 48-60, 2011.

[14] Poonkuzhali, Karthiprakash, Valarmathy and kalamani, "An Approach To Feature Selection Algorithm Based On Ant Colony Optimization For Automatic Speech Recognition", *Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 2, No. 11, pp. 5671-5678, 2013.

[15] I. E. Henawy, W. Khedr, O. ELkomy and A. Z. Abdalla, "Recognition of phonetic Arabic figures via wavelet based Mel Frequency Cepstrum using HMMs", *Journal of Housing and Building National Research Center*, Vol. 10, pp. 49-54, 2014.