International Journal of Advanced Trends in Computer Applications
www.ijatca.com

REVIEW OF SPEECH AND SPEECH RECOGNITION SYSTEM USING FEATURE EXTRACTION ALGORITHM AND OPTIMIZATION ALGORITHMS

[1]Gurpreet Kaur, [2] Dr. Ravinder Khanna, [3] Dr. Amod Kumar
[1]Assistant Professor, [2] Professor, [3] Chief Scientist
[1]UIET, PU, Chd.,[2] Maharishi Markandeshwar University, Ambala,[3]BIU, CSIR-CSIO, Chd.

ABSTRACT--This paper analyses the audio inconsistency of speakers and its impact on the strength of existing automatic speech recognition and speaker recognition systems. The acoustic and visual features are evaluated by a Support Vector Machine for digit and speaker detection and later by Hidden Markov Model verification. A methodology for speech recognition with speaker recognition based on Hidden Markov Model for security is a requirement of science. Mapping of speech using Artificial Neural networks is obtainable. Fireflies create glowing flash as a sign scheme to correspond with additional fireflies particularly to prey attractions. Cuckoo search algorithm is used as its search space is extensive in nature. Genetic algorithm can shun calculating system slope in traditional gap investigation and determines the optimum interval range of the parameters under acceptable corresponding aim error boundary. In order to obtain to obtain the most efficient and linearly discriminative components, LDA is used.

Index terms—Speech Recognition System, LDA, Neural Network, Genetic Algorithm, SVM, Optimization Algorithms, Cuckoo Search.

I. INTRODUCTION

In today's era, people are more contented in Speech, accordingly they like to interact with computers with speech, except using primordial interfaces like keyboards and pointing devices. Automatic Speech Recognition is the system that helps to recognize the words that a person speaks via telephones, microphone or any hearing device and then convert that to written form. As communication is the only way with which the person share their views/ideas but it is dominated on language basis. Therefore, computer could be the only way out with which the recognition of the native language takes place. In speech recognition, the engine that works to handle complex tasks those is taking input in form of audio and convert that to a text that a user / application can understand [8-9].

There are basically two types of speech data: Speaker dependent and speaker independent. In speaker dependent systems, the system is inputted with every person's voice for good accuracy, which has been achieved by training phase. The main advantage of speaker dependent systems is that it has less error rate w.r.t speaker independent systems i.e. 3 or 4 times better. So in proposed work speaker dependent system development will be done. The speaker recognition is a very challenging task because lots of variables are needed in this work like pitch, bandwidth, frequency, spectral features, amplitude etc. In other words we can say that when speaker is known to listener, it is also called speaker dependent. There is lot of work that has been done in this context but still very much enhancement is need so in this paper various algorithms will be reviewed in next section like HMM, SVM, Firefly algorithm.

1.1 Mathematical representation of speech signal

B (y, u) is the time varying variable, r (u) is the non-stationary change in time. B (y, u) is changing w.r.t r (u) but in slow time constant and can be represented as:

$$\lfloor r(k) \rfloor = r(k\Delta\,\omega)$$
$$= H(k(k\Delta\,\omega) + Zr((k\Delta\,\omega)$$
$$\frac{\partial A}{\partial t} \ll \frac{\partial p}{\partial t}$$

Where p(t) is sequence of intervals.

Four components with which the recognized text can be achieved as shown by the figure below:

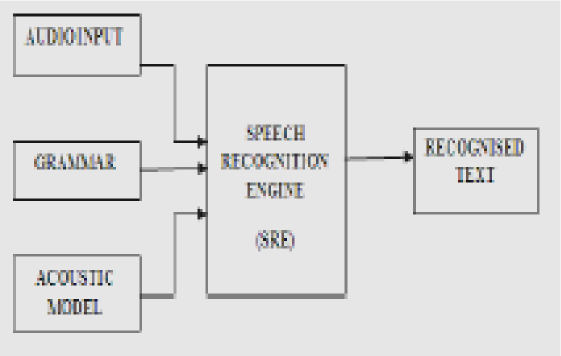


Figure 1: Components of Speech Recognition

The audio input /audio stream is very immaculate, it comes into the recognition engine and it consists of background noise, speech data. The speech engine has to switch to the situation within which the audio is spoken. The SRE needs all types of data, algorithms and statistics and then convert that to the format which can be sustained further [10]. After the speech data describes it best format, the exact match 1for the engine that suits it can be searched by taking the words, phrases that it knows with the environment knowledge in which the operating can be done, takes into consideration. Acoustic model gives the knowledge of the environment needed. It returns the value when the text string can be recognized.

1.2 TYPES OF SPEECH

SRS (Speech Recognition System) can be separated in various classes by describing what type of utterances they can distinguish.

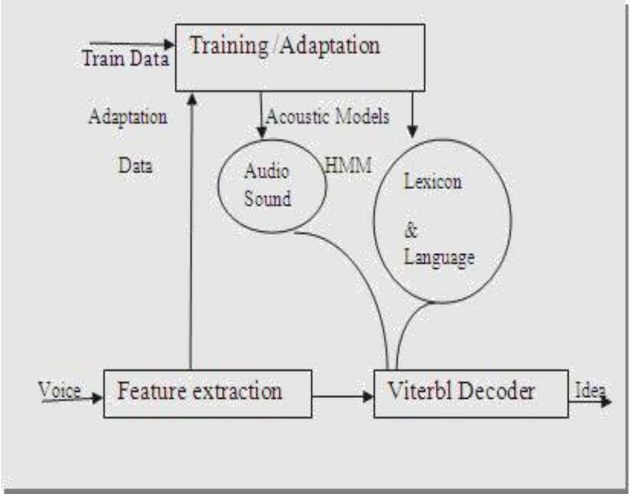


Figure 2: Speech recognition System

LITERATURE REVIEW

Research Gap:

Emotional Speech recognition is not an easy task. Earlier works which have been done on this particular area used Mel Frequency Cepstral Coefficients (MFCC) [11, 12], pitch frequencies as formants [14, 15] speech rate [16], energy [19,12,18], linear predictive coding (LPC) [21] for feature extraction purposes. In addition various classification techniques that are used in emotion classification task contains Support Vector Machines (SVM) [2,3], Neural Networks (NN) [10], Hidden Markov Models (HMM) [12] and Linear Discriminant Analysis (LDA) [23,24]. In proposed work we will try to use above mentioned techniques in hybridization and using advanced features of speech like bandwidth, frequency, pitch ratio etc.

Already Work done:

Shashidhar et.al [25], presented the recent literature on speech emotion recognition with considering the issues related to emotional speech corpora, different types of speech features and models used for recognition of emotions from speech. Chengwei et al. [26], Study the cross-database speech emotion recognition based on online learning. How to apply a classifier trained on acted data to naturalistic data, such as elicited data, remains a major challenge in today’s speech emotion recognition system. Vipul Garg et al.[27], discusses a novel design/approach based on a hierarchical decision tree for the GMM means super vector based feature set and using various classifiers viz. SVM, BLG and SVR, to improve the performance of the existing emotion recognition systems. Arianna et al. [28], propose the use of a PLS regression model, optimized according to specific features selection procedures and trained on the Italian speech corpus EMOVO, suggesting a way to automatically label the corpus in terms of arousal and valence.

II. BASIC TECHNOLOGY

Basic technology includes a pattern recognition paradigm which is based on HMM (Hidden Markov Model) as shown below:

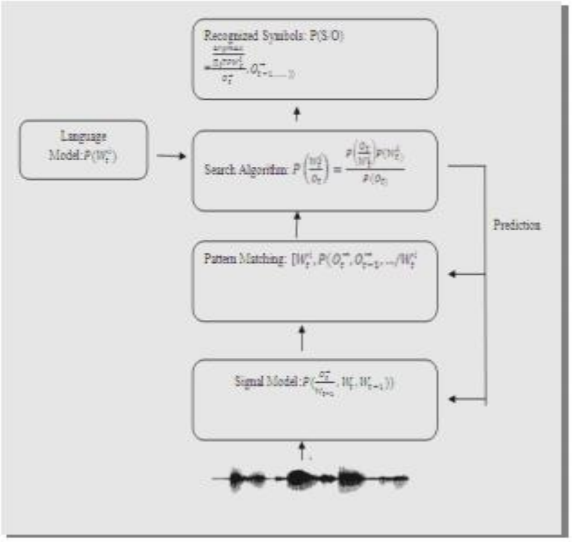


Figure 3: Basic technology of SRS

2.1 SPEECH CODING

The main aim of speech coding is to get savings in required memory storage space, transmission bandwidth and transmission power by reducing the number of bits per sample as the decoded speech is perceptually identical from the unique speech. Speech coding methods can have the following gains [12]:

- Reductions in speech rate and memory storage are directly proportional to bit rate.
- Reduction in transmission power need as after compression, less number of bits per second to transmit is required.
- Error control coding methods can be used to introduce saved bits per sample again to protect speech Para-metrics from channel noise as noise immunity.

III. FEATURE EXTRACTION ALGORITHM

3.1 MFCC (Mel Frequency Cepstral Coefficients)

The most popular spectral based parameter used in recognition approach is the Mel Frequency Cepstral Coefficients called MFCC. Due to its advantage of less complexity in implementation of feature extraction algorithm, only sixteen coefficients of MFCC corresponding to the Mel scale frequencies of speech Cestrum are extracted from spoken word samples in database. All extracted MFCC samples are then statistically analyzed for principal components, at least two dimensions minimally required in further recognition performance evaluation.

The Mel scale relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. Humans are much better at discerning small changes in pitch at low frequencies than they are at high frequencies. Incorporating this scale makes our features match more closely what humans hear. The formula for converting from frequency to Mel scale is:

$$M(f) = 1125\, In \left(l + \frac{f}{700}\right)$$

To go from Mels back to frequency:

$$M^{-1}(m) = 700(\exp\left(\frac{m}{1125}\right) - 1)$$

The major stages of MFCC can be summarized as follows:

A. Pre-emphasis

A pre-emphasis of high frequencies is therefore required to obtain similar amplitude for all formants. Such processing is usually obtained by filtering the speech signal with a first order FIR filter whose transfer function in the z-domain is:

$$H(z) = 1 - \propto . z - 1$$

α, being the pre-emphasis parameter. In essence, in the time domain, the pre-emphasized signal is related to the input signal.

B. Windowing

Traditional methods for spectral evaluation are reliable in the case of a stationary signal. For voice, this holds only within the short time analysis can be performed by "windowing" a signal x'(n) into a succession of windowed sequences xt(n) t=1, 2,…,T, called frames, which are then individually processed:

$$x^{'}(n) = x^{'}(n - 1.Q), 0 \leq n < N1 \leq t \leq T(3)$$
$$xt(n) = w(n).x'(n)$$

where, w(n) is the impulse response of the window. Each frame is shifted by a temporal length Q. If Q=N, frames do not temporally overlap while if Q<N, N–Q samples at the end of a frame x'(n) duplicated at the end of the following:

$$frame\, x't + 1(n)$$

C. Spectral analysis

The standard methods for spectral analysis rely on the Fourier transform of xt (n): Xt (ejω).Computational complexity is greatly reduced if Xt (ejω) is evaluated only for a discrete number of ω values. If such values are equally spaced, for instance considering $\omega = \frac{2\pi k}{N}$, then the discrete Fourier Transform (DFT) of all frames of the signal is obtained:

$$X1(k) = Xt\left(\frac{ej2\pi k}{N}\right), k = 0, \dots, , N - 1 (4)$$

D .Filter Bank Processing

Spectral analysis reveals those speech signal futures which are mainly due t the shape of the vocal tract. Spectral futures of speech are generally obtained as the exit of filter banks, which properly integrate a spectrum at defined frequency, ranges [11]. A set of 24 band-pass filters is generally used since it simulates human ear processing. There are many methods to implement such filters. A computationally inexpensive method consists of performing filtering directly in the DFT domain.

E. Log Energy Computation

The previous procedure has the role of shooting the spectrum, performing a processing that the similar to that is similar to that executed by human ear [11]. The next step consists of computing the algorithm of square magnitude of the coefficients Y1(m) obtained with above eq. This reduces the simply computing the logarithm of magnitude of the coefficients, because of logarithm algebraic property which brings back the logarithm of a power to a multiplication by a scaling factor. The final procedure for the Mel frequency cestrum computation (MFCC) consists of performing the inverse DFT on the logarithm of the magnitude of the filter bank output [12].

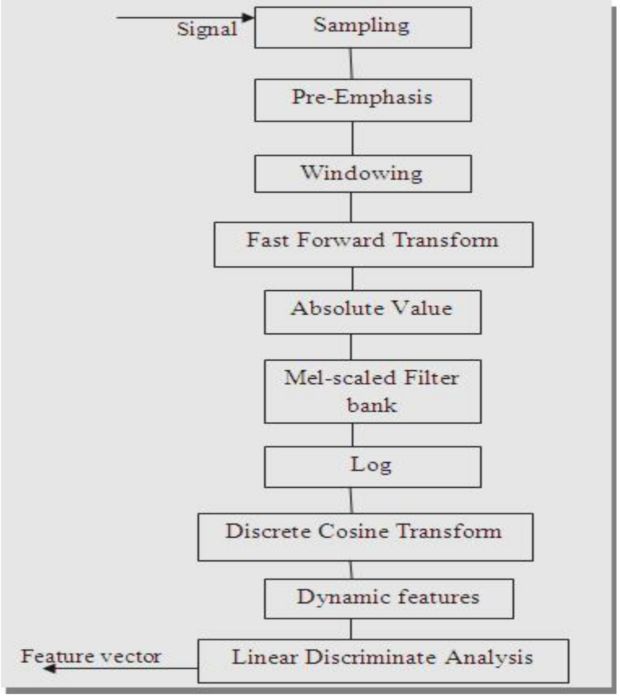


Figure 4: MFCC Process

3.2 GTCC

GTCC algorithm is used for characteristic extraction. It include Fast Fourier revolution used to convert the time domain signal to frequency domain for phantom analysis, filtration development like hamming window which is a type of filter to attenuate the unwanted frequencies and accepts the required frequency to boost up the frequencies and Error rectangular bandwidth which is the process of bandwidth approximation and to increase the strength of the signal in noisy environment.

3.3 LPCC

LPCC analysis is an effective method to estimate the main parameters of speech signals. The conclusion extracted was that an all-pole filter, H(z) , is a good approximation to estimate the speech signals. Its transfer function was described. In this way, from the filter parameters (coefficients, {ai}; and gain, G), the speech samples could be synthesized by a difference equation. Thus, the speech signals resulting can be seen as linear combination of the previous p samples. Therefore, the speech production model can be often called linear prediction model, or the autoregressive model. From here, p, indicates the order of the LPC analysis; and, the excitation signal, e[n], of the speech production model can be called prediction error signal or residual signal for LPC analysis.

IV. GENETIC ALGORITHM

According to Goldberg et al., 1989, GA is commonly used in applications where search space is huge and the precise results are not very important. The advantage of a GA is that the process is completely automatic and avoids local minima. The main components of GA are: crossover, mutation, and a fitness function. A chromosome represents a solution in GA [19]. The crossover operations used to generate a new chromosome from a set of parents while the mutation operator adds variation. The fitness function evaluates a chromosome based on predefined criteria. A better fitness value of a chromosome increases its survival chance. A population is a collection of chromosomes. A new population is obtained using standard genetic operations such as single-point crossover, mutation, and selection operator. As a GA is relatively computation intensive, this chapter proposes executing the algorithm only at the base station. The proposed GA is used to generate balanced and energy efficient data aggregation trees for wireless sensor networks. The following sections present the design of the proposed GA.

Population: A population is a collection of chromosomes. A new population is generated in two ways: steady state GA and generational GA. In steady-state GA, a few chromosomes are replaced in each generation; whereas the generational GA replaces all the current chromosomes in each generation. A variation of generational GA is called elitism, which copies a certain number of the best chromosomes from the current generation to the new generation. A new generation is created using crossover and mutation operations. In case of the proposed GA, a population is a collection of possible aggregation trees. For the initial population, the parent nodes are selected arbitrarily and the validity of chromosomes is also maintained. The size of the population remains fixed for a specific experiment and remains the same for all generations.

The proposed GA uses the elitism technique to retain the best chromosome in each generation. Further, the rest of the chromosomes are replaced by using crossover and mutation operations.

Fitness: According to Kreinovich et al., 1933, in nature, fitness of an individual is the ability to pass on its genetic material.

This ability includes an individual's quality to survive. In GA, the fitness of a chromosome is evaluated using a defined function to solve a problem. A chromosome with a higher value has the better chance of survival.

Selection: The selection process determines which chromosomes will mate (crossover) to produce a new chromosome. A chromosome with a higher fitness value has a better chance of selection for mating. Several selection method sexist to select chromosomes for mating such as: "Roulette-Wheel" selection," Rank" selection and "Tournament" selection. This chapter uses

Tournaments election, where a pair of chromosomes is chosen randomly from the current population. Between these two chosen chromosomes, the fit chromosome is selected with a predefined probability p; whereas the other chromosome is selected for mating with the probability (1-p), as described by Goldberg et al., 1989. Crossover is known as a binary genetic operator acting on a pair of chromosomes. Crossover recombines genetic material of two participating chromosomes. It simulates the transfer of genetic inheritance during the sexual reproductive process. The crossover result depends on the selection process made from the population. We use a crossover operation where the crossover point is randomly selected and the gene values of participating parents are flipped to create a pair of child chromosomes. The pair of children is produced by flipping the gene values of the parents after the cross over point. In this example, the last two genes are flipped.

Repair Function. Notice that the crossover operation between two valid chromosomes may produce invalid chromosome(s). A repair function is

```
procedureRepairFunction(chromosome)

for each gene iteration in chromosome do
    while geneiteration creates a cycle do
        randomly select a new parent for gene
    end while
end for
```

The mutation adds variation in the next generation. In mutation, a node is randomly picked and its parent ID is reselected randomly. Similar to crossover, the mutation operation may produce an invalid chromosome, which is also fixed using the repair function.

Fitness Parameters The fitness parameters are designed with two objectives. First, we need an energy efficient aggregation tree so that nodes can communicate for a longer period of time. Second, the generated tree should balance load among nodes. A few fitness parameters are described in this section. The load of a sensor node represents the communication energy required in a single round. A sensor node has different loads in different aggregation trees. The tree generated by the GA needs to be energy efficient. It is preferable that each node spends less amounts of energy in a communication round. For a node i, the term $\sum Ei$ denotes the energy consumption ratio in the tree. We want to maximize this ratio for all nodes.

Step 1 : At random, produce an initial population M (0).

Step 2 : Compute as well as help save the actual fitness f(m) for every specific individual m in the current population M(t);

Step 3: Specify selection probabilities p (m) for every specific individual m throughout M (t) making sure that p (m) is actually proportional to f(m).

Step 3 : Specify selection probabilities p (m) for every specific individual m throughout M (t) making sure that p (m) is actually proportional to f (m).

Step 4 : Crank out M (t+1) by simply probabilistically choosing individuals from M (t) to produce offspring via genetic operators.

Step 5: Repeat step 2 until satisfying solution is actually attained.

However, in order for genetic algorithms to work effectively, a few criteria must be met [20]:

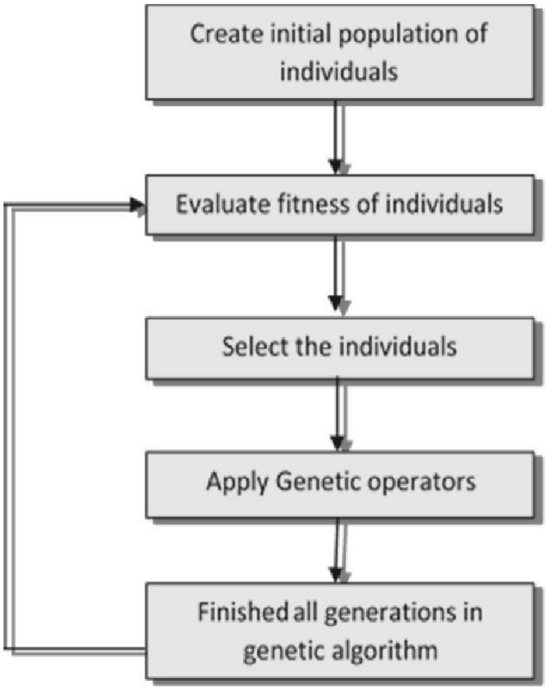


Figure 5: Flow Chart of GA

- It must be relatively easy to evaluate how "good" a potential solution is relative to other potential solutions.
- It must be possible to break a potential solution into discrete parts that can vary independently. These parts become the "genes" in the genetic algorithm.
- Finally, genetic algorithms are best suited for situations where a "good" answer will suffice, even if it's not the absolute best answer.

Population Size	Here the extracted key points through SIFT algorithm will be considered as Population Size
Mutation	**Mutation is the change in State**
Cross Over	**Linear**

V. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) also called Support Vector Networks are supervised learning models that analyze data and recognize patterns. SVM models represent examples as point in space mapped in manner that separate categories examples are divided by a gap thereby performing linear classification. Apart from this SVMs can also perform nonlinear classification using Kernel trick [2] [3].

SVM was developed by Vapnik in 1998 and it is a new class of learning machine which use support vectors and kernels for learning. The kernel machines provide a framework which can be adapted to different tasks and domains by choosing appropriate kernel function. Basically, SVM was developed for binary classification and later on it was extended to solve multi class classification using one against all strategy. In multi class SVM, multiclass labels are decomposed into several two class labels and it trains the classifiers to solve the problems and the solution of multi class problem is reconstructed from the outputs of classifiers. The one against all strategy comprises of constructing one SVM per class, which is trained to distinguish the samples of one class from the samples of all remaining classes. SVMs are very widespread apprentice. In their fundamental form, SVMs study linear threshold function. Support vector machines are based on the Structural Risk Minimization theory from computational knowledge hypothesis. SVM are independent of the dimensionality of the feature space. Characteristics of SVM:

High dimensional input space

- Document vectors are sparse
- Few irrelevant features
- Mainly text classification problems are linear.

The main idea of SVM is that; it finds the optimal separating hyper plane such that error for unseen patterns is minimized. Consider the problem of separating the set of training vectors belonging to two separate classes.

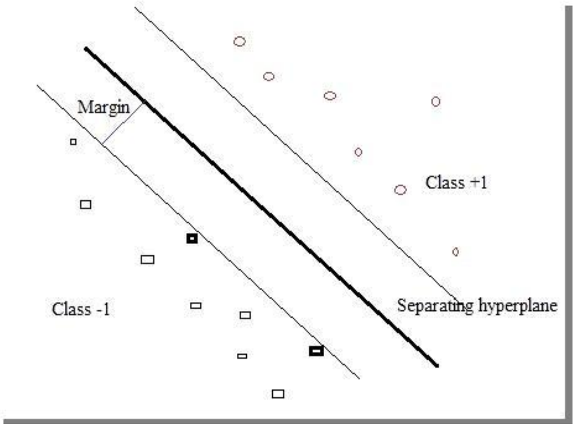


Figure 6: SVM Working

$$x_1, x_2, \dots . x_n$$

Which are vector in RD,
We consider a decision function of the following form [14]:

$$yx = wT\emptyset x + b$$

Attached to each observation, $x_,$ is a class label, $t_i \in -1, +1$. Without loss of generality, we must construct a decision function such that, $yx_i > 0$ for all i such that $t_i = +1$ and $yx_i < 0$ for all I such that $t_i = -1$. We can combine these requirements by stating,

$$t_i yx_i > 0 \forall i$$

The idea is to extend it to multi-class problem is to decompose an M-class problem into a series of two-class problems.

VI. BACTERIAL FORAGING OPTIMIZATION (BFO)

BFO algorithm is first projected by Passino in 2002 [6]. It is motivated by the foraging and Chemo tactic behaviors of bacteria, especially the Escherichia coli (E. coli). During foraging of the real bacteria, locomotion is achieved by a set of tensile flagella. Flagella help an E.coli bacterium to fall or swim, that are two essential operations performed by a bacterium at the instance of foraging. When they revolve the flagella in the clockwise direction, every flagellum pulls over the cell. That results in the moving of flagella separately and lastly the bacterium tumbles with smaller amount of tumbling while in a damaging place it tumbles repeatedly to find a nutrient gradient. Stirring the flagella in the counterclockwise direction helps the bacterium to swim at a very speedy rate. In the above mentioned algorithm the bacteria undergoes chemo taxis, where they like to shift towards a nutrient gradient and shun harmful atmosphere. Usually the bacteria shift for a longer distance in a gracious situation. BFO is a novel class for biological stochastic global search method on copying the foraging the behavior of E.coli bacteria and it is useable for locating, handling and digesting the food. While foraging, the bacterium contains two types of actions that are: Tumbling and Swimming. The tumble modifies the beginning of bacterium and while swimming, the bacterium shifts in its original position. The The tumble modifies the beginning of bacterium and while swimming, the bacterium shifts in its original position. The **Step1:** Initialize parameters p,

Step 1: $N_c, N_s, N_r, N_{ed}, P_{ed}$ C(i)(i=1,2,…S), θ^i.
Step 2: Elimination dispersal loop:l=l+1
Step 3: Reproduction loop;k=k+1.

Step 4: Chemo taxis loop:j=j+1.

[a] For i=1,2,…S take a chemotatic step for bacterium I as follows.

[b] Co*mpute fitness function (i,j,k,l).

Let ,J(i,j,k,l) = J(i,j,k,l)+ J_{cc} (θ^i (j,k,l),P(j,k,l).

[c] Let J_{last}= j(i,j,k,l) to save this value since we may find a better cost via a run.

[d] Tumble: generate a random vector (i) $\in$ R^p each element $_m$(i),m=1,2,…,p, a random number on [-1,1].

[e] Move: Let

$$\theta^i(j+1,k+1) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

This results in a step od=f size C (i) in the direction of the tumble for Bacterioum i.

$$J(i,j+1,k,l) = J(i,j,k,l) + j_{cc}\theta^i(j+1,k.l),P(j+1,k,l))$$

[f] Compute J (I,j+1,k,l) and let

$$J(i,j+1,k,l) = J(i,j,k,l) + j_{cc}\theta^i(j+1,k.l),P(j+1,k,l))$$

[g] Swim

i. Let m=0
ii. While m<N_s
    - Let m=m+1
    - If j(I,j+1,k,l)<J_{last} ,let J_{last}=J(I,j+1,k,l) let,

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

And use this $\theta^i(j+1,j,k)$ to compute the new $J(i,j+1,k,l)^i$ as we did in [f].

iii. Else, let m=N_s. This is the end of the while statement.

[h].Go to next Bacterioum (i+1) if i $\neq$S.
Step 5: If j<N_c , go to step 4.is

Step 6: Reproduction:
[a]. For the given k and l , and for each i=1,2,..S.let

$$J^{i}_{health} = \sum_{j=1}^{N_c+1} J(i, j, k, l)$$

Be the health of bacterium i. Sort bacteria and Chemo tactic parameters C(i) in order of ascending cost J_{health}.

[b]. The S_r bacteria with the highest J_{health} values die and the remaining S_r bacteria with the best values split.

Step 7: if $k<N_{re}$, go to step 3.
Step 8: Elimination-dispersal: For i=1,2,…S with probability P_{ed}, eliminate and disperse each bacterium. To do this, if a bacterium is eliminated, simply disperse another one to a random location on the optimization domain. If $l< N_{ed}$, then go to step 2; otherwise end

VII. CUCKOO SEARCH

It originates from the behavior of number of species of cuckoo that lay eggs in another birds nest in parasitic way. It would live if the cuckoo eggs adopt the behavior of host nests or else it would discard the cuckoo's egg. Algorithm for cuckoo search is described below [22]:

- Comparison is done between cuckoo's egg and set of host nests.
- Levy flight is used to bring in uncertainty to decide host nests.
- Two sets of solution that is quality solution and discarded solution are compared.
- Best solution can be getting by ranking function through quality solution and the worst nests are detached with the suitable probability.

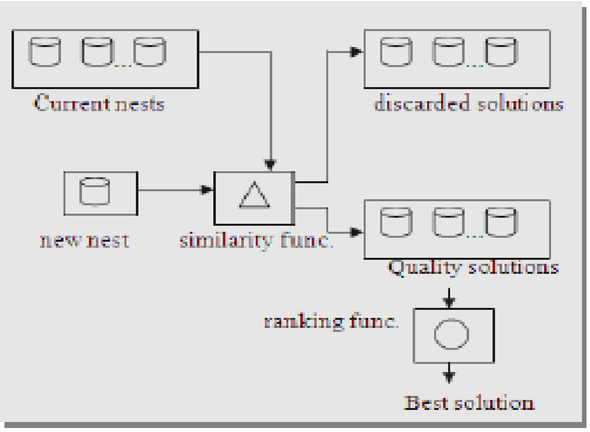


Figure 7: Cuckoo Search Procedure

VIII. NEURAL NETWORK

Machine learning algorithms help a lot in decision making and neural network has performed well in classification purpose in medical field. Most popular techniques among them are neural network. Neural networks are those networks that are composed of simple elements which operate parallel. A neural network [10] can be trained to perform a particular function by adjusting the values of the weights between elements. Network function is determined by the connections between elements. There are several activation functions that are used to produce relevant output. Algorithm for neural network is as follows:

The algorithm is as follows:

(1) Step 0. Initialize weights. (Set to small random values)
(2) Step 1. While stopping condition is false do steps 2–9
(3) Step 2. For each training pair do steps 3– 8.

A. Feed-forward

Steps 3: Each input unit (Xi, i= 1 . . . n) receives input signal Xi and sends signal to all units in the layer above (the unseeable units).

Steps 4: Each hidden unit (Zj, j = 1, p) sums its weighted input Signals.

V0j: Bias on hidden unit j.

Vij : Weight between input unit and hidden unit. It applies its activation function to calculate its output signal, and convey that signal to all units in the layer above (output units).

Steps 5: Each output unit (Yk, k = 1 . . . m) sums its burdened input signals

Wok: Bias on output unit k.

Wjk: Weight between hidden unit and output unit.

And uses there activation function to compute its output signal.

B. Back propagation of error

Step 6: Each output unit (Y_k, k= 1. . . m) receives a target pattern matching to the input training pattern, calculates its error information term.

Step 7: Each hidden unit (z_j, j = 1 . . . p) adds up its delta inputs. To multiply the derivative of its activation function to calculate its error information term and calculates its bias correction term.

Step 8: Each output unit (Yk, k = 1 . . . m) updates its bias and weights (j = . . . p)

Unit (Zj, j = 1. . . p) updates its bias and weights (i= 0, . . . , n):

$$Vij(new) = Vij(old) + \Delta\, Vij$$

Step 9: Test stopping condition.

IX. FUZZY SUPPORT VECTOR MACHINES (SVM)

When dealing with standard SVMs, all training vectors are presumed to belong entirely to either class +1 or class −1, and moreover are assumed to have equal weight or relevance. However in many applications this assumption may not hold, as some training vectors may have more importance than others, and sometimes we may not be entirely certain that the purported classification d_i of a training vector xi is actually correct [4-7]. Fuzzy theory deals with these issues by saying that a training vector xi belongs, for example, 90% to class +1 and 10% to class −1. This may be achieved by associating a fuzzy membership $0 < s_i \leq 1$ with each training pair (x_i, d_i) . The pair is then regarded as having a membership of s_i to class s_i; and a membership $1 - s_i$ to class − (although the latter is usually ignored in FSVM theory). This may be incorporated into the SVM framework by modifying the primal training problem.

X. FIREFLY ALGORITHM

The Firefly algorithm is a freshly developed nature-inspired Meta heuristic algorithm [22]. The Firefly algorithm is encouraged by the social presentation of fireflies. Fireflies may also be called lightning bugs. There are about 2000 firefly species in the globe. Most of the firefly species construct short and rhythmic flashes. The model of flashes is unique for a particular species. A firefly's twinkle mainly acts as a signal to attract mate partners and potential prey. Flashes also serve as a defensive warning instrument. The following three idealized rules are considered to explain the firefly algorithm [23]:

1) All fireflies are unisex so that one firefly will be involved to other fireflies despite of their sex.
2) Attractiveness is relative to their brightness; thus, for any two flashing fireflies, the less bright one will move in the direction of the brighter one. The attractiveness is relative to the brightness and they both reduce as their distance increases. If there is no brighter one than a particular firefly, it will move arbitrarily.

The clarity of a firefly is affected or unwavering by the landscape of the idea function. For a maximization problem, the brightness may be comparative to the objective function value. For the minimization problem, the brightness may be the give-and-take of the objective function value. The make believe code of the firefly algorithm was given by Yang.
The pseudo code of the firefly algorithm is known in Algorithm 1.

A. Attractiveness

The attractiveness of a firefly is determined by its light intensity. The attractiveness may be calculated by using the equation:

$$\beta(r) = \beta_o e^{-r^2}$$

B. Distance

The distance among any two firefly's k and l at X_k and X_l is the Cartesian distance as follows

$$r_{ld} = \|x_k - x_l = \sqrt{\sum_{k=1}^{d} (x_{k,o} - x_{l,o})^2}$$

C. Movement

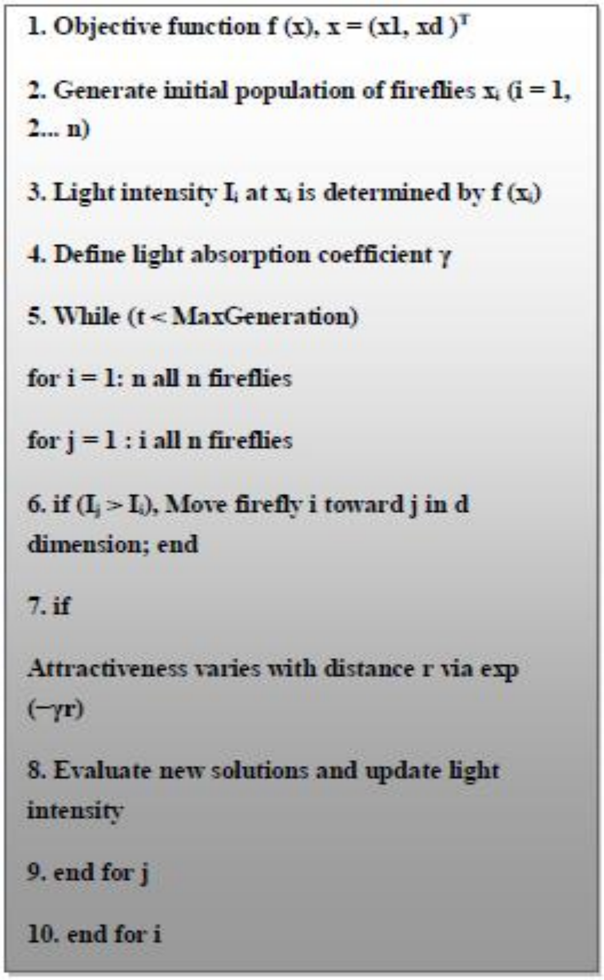
The movement of a firefly *k* that is attracted to another more attractive firefly *l* is determined by 0.

$$x_k = x_k + \beta_o e^{-r^2} (X_l - X_k) + \alpha(rand - \frac{1}{2}$$

XI. LDA (Linear Discriminate Analysis)

LDA seeks directions that are efficient for discrimination between classes. This method is used in statistics, pattern recognition and machine learning to locate a linear combination of features that characterizes or separates two or more classes of objects or events. The resultant mixture might be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification [23] [24].

Pseudo code of the Firefly Algorithm



```
1. Objective function f (x), x = (x1, xd )^T
2. Generate initial population of fireflies x_i (i = 1, 2... n)
3. Light intensity I_i at x_i is determined by f (x_i)
4. Define light absorption coefficient γ
5. While (t < MaxGeneration)
    for i = 1: n all n fireflies
        for j = 1 : i all n fireflies
            6. if (I_j > I_i), Move firefly i toward j in d dimension; end
            7. if
            Attractiveness varies with distance r via exp (−γr)
            8. Evaluate new solutions and update light intensity
        9. end for j
    10. end for i
```

For iteration k for data U_k:

$$\mu_{xj} = \frac{\sum_{class\,j} Z_k}{n_j} = A\frac{\sum_{class\,j} Z_k}{n_j} = A\sum_{class\,j}(X_k-\mu) = A\mu_j - \mu$$

And the general formula:

$$(D+BC)^{-1} = D^{-1}-D^{-1}B\,(I+CD^{-1}B)^{-1}CD^{-1}$$

The updated formula S_w is given by:

$$S_w = S_w + (Z_k-\mu_{kj})(Z_k-\mu_{kj})^T = S_w + A(Y_k-(\mu_j-\mu))(Y_k-(\mu_j-\mu))^T A^T$$

XII. CONCLUSION

In recent years, the requirement for speech recognition research based on large vocabulary speaker independent permanent speech has highly enlarged. The fundamentals of speech recognition are discussed and its recent progress is investigated. In this paper, we explored the core components of modern statistically based speech recognition systems, Cuckoo search for randomness, NN is also presented for the validation, FA for flashing behavior of fireflies, Genetic Algorithm to remove the security risks and power utilization of the system. LDA for finding the guidelines that will well split the diverse classes of the data once projected upon. SVM applied to network intrusion detection system.

REFERENCES

[1]. Li Kun-Lun, HUANG Hou-Kuan, Fuzzy Multi-Class support vector machine and application in instrusion detection, Chinese Journal Of Computers Vol. 28 No.5 Feb. 2005.

[2]. XiaoLi Lv, Lei Li, Research on the Methods of SVM Multi-class Classification Based on Binary Tree in the Face Recognition System.

[3].HongFei Lin WeiHua Zhuang HoHai University Research of support vector machine and its application in intrusion detection.

[4]. Takuya Inoue, "Fuzzy Support Vector Machines for Pattern Classification", IEEE, pp. 1449-1454, 2001.

[5]. Joanna Czajkowska, Marcin Rudzki, and Zbigniew Czajkowski, "A New Fuzzy Support Vectors Machine for Biomedical Classification", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, pp. 4476-4479, August 20-24,2008.

[6]. Xuehua Li, Lan Shu, "Fuzzy Theory Based Support Vector Machine Classifier", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, pp. 600-604, 2008.

[7]. Bai-Xing Ii, Qian Xu, "Medical Image Classification based on Fuzzy Support Vector Machines", 2008International Conference on Intelligent Computation Technology and Automation, pp. 145-149,2008.

[8]. Niladri Sekhar Dey, Ramakanta Mohanty , “Speech and Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model,” 2012 International Conference on Communication Systems and Network Technologies 978-0-7695-4692-6/12 $26.00 © 2012 IEEE DOI 10.1109/CSNT.2012.221

[9]. L. R. Bahl et al., “Speech Recognition of a Natural Text Read as Isolated Words,” Proc. IEEE Int’I Conf. Acoustics, Speech, and Signal Processing, April 1981, pp. 1,168-1,17 1.

[10]. D. O’Shaughnessy, Speech Communication: Human and Machine, Addison- Wesley, Reading, Mass., 1987.

[11]. J. Mariani, “Recent Advances in Speech Processing,” Proc. IEEE Int’l Conf. Acoust i c s , Speech, and Signal Processing, Glasgow, Scotland, May 1989, pp. 429- 440.

[12]. H. Murveit and M. Weintraub, “1,000- Word Speaker-Independent Continuous- Speech Recognition Using Hidden Markov Models,” Proc. IEEE Int’l Conf, Acoustics, Speech, and Signal Processing, New York City, 1988, pp. 115-1 18.

[13]. J.D. Markel and B. Davis, “Text-Independent Speaker Recognition from a Large Linguistically Unconstrained Time-Spaced Database,” IEEE Trans. iiASSP-27, No. I , 1979, pp.

74-82.

[14]. C.-C. Chang and C.-J. Lin, “LIBSVM—A Library for Support Vector Machines, ”www.csie.ntu.edu.tw/cjlin/libsv m, 2001.

[15]. M. Schmidt and H. Gish, “Speaker Identification via Support Vector Classifiers,” Proc. IEEE Int’l Conf. Acoustics, Speech, and Signal Processing (ICASSP ’96), pp. 105-108, 1996.

[16]. Wasserman, P.D. Neural Computing: Theory and Practice. New York: Van Nostrand Reinhold, 1989.

[17]. Stanley, J. Introduction to Neural Networks. Sierra Madre, CA: California Scientific Software, 19X9.

[18]. Hopfield, J.J. and Tank. D. “Neural Computation of Decisions in Optimization Problems”, Biologicnl Cyhrrnetics, Vol. 52, 1985. pp. 147-152.

[19]. Guo Pengfei, Han Yingshi, “Chaotic genetic algorithm for structural optimization with discrete variables”, Journal of Liaoning Technical University, 2007, 26(1), pp. 68-70.

[20]. Pengfei Guo Xuezhi Wang, “The Enhanced Genetic Algorithms for the Optimization Design,” 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010) 978-1-4244-6498-2/10/$26.00 ©2010 IEEE.

[21]. Xin-She Yang, Suash Deb. “Cuckoo Search via levy flights,” In proceedings of world congress on nature & biologically inspired computing , IEEE publications USA, pp 210-214, Dec.2009,India.

[22]. L.Cui and H.Wang, “Reach back Firefly synchronicity with Late Sensitivity Window in Wireless Sensor Networks,” Proc. of IEEE Ninth International Conference on Hybrid Intelligent Systems, pp. 451-456, 2009.

[23]. A. M. Martinez and A. C. Kak, “PCA versus LDA”, IEEE Trans. Pattern Anal. Mach. Intell. vol. 23, no. 2, pp. 228–233, Feb 2001.

[24]. Yu, H., Yang, J.: A direct LDA algorithm for high dimensional data with application to face recognition. Pattern Recognition, Vol. 34. No. 11. (2001) 2067-2070.

[23]. Shashidhar G. Koolagudi · K. Sreenivasa Rao, “Emotion recognition from speech: a review”, Springer, Vol.15 pp. 99-117, 2012.

[24]. Yu, H., Yang, J.: A direct LDA algorithm for high dimensional data with application to face recognition. Pattern Recognition, Vol. 34. No. 11. (2001) 2067-2070.

[25]. Shashidhar G. Koolagudi · K. Sreenivasa Rao, “Emotion recognition from speech: a review”, Springer, Vol.15 pp. 99-117, 2012.

[26]. Chengwei Huang, Ruiyu Liang, Qingyun Wang, Ji Xi, Cheng Zha, and Li Zhao, “Practical Speech Emotion Recognition Based on Online Learning: From Acted Data to Elicited Data”, Hindawi: Mathematical Problems in Engineering, pp. 1-9, 2013.

[27]. Vipul, Harsh and Rohit , “Speech Based Emotion Recognition Based on Hierarchical Decision Tree with SVM, BLG and SVR Classifiers”, IEEE, Vol.978 , pp. 4673-5952

[28]. Arianna Mencattini , Eugenio Martinelli , Giovanni Costantini, Massimiliano Todisco Barbara Basile, Marco Bozzali, Corrado Di Natale, “Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure”, ELSEVIER, Online, ISSN : .2014.03.0192014, 2014.