CHAPTER 3

SELECTION OF COMBINATION OF FEATURES USING GENETIC ALGORITHM

3.1 INTRODUCTION

Feature selection is an important task in any pattern recognition application. Selecting a useful and relevant subset of features from a large set is important to significantly improve the performance of recognition system. Feature selection has become important in areas such as bioinformatics (Basiri et al 2008; Nemati et al 2009), signal processing (Liao 2010; Roda 2008), image processing (Tsai et al 2012; Choi et al 2012), and speaker verification (Nemati et al 2011). The basic concept of feature selection in speech processing is to select the features which are most important to represent the speaker and eliminating the less important features from further processing. This reduces computation time and storage requirements in real time implementation. To select an appropriate set of features, a criterion function can be used to provide the discriminatory power of the individual features.

Feature selection algorithms are classified into two categories based on 1) the processes involved and 2) the method of modelling (Guyon & Elisseeff 2003). Based on method of modelling, the categories are model free methods and model based methods. In model free method, statistical analysis of data is performed for a particular model. Whereas in the model based method, the different sets of features are applied to models and the performance of the model for various features are compared. The set of
features which produces the minimum error or maximum performance is selected. Based on the processes involved, the feature selection categories are feature ranking method and subset selection method. In feature ranking method, the features are ranked by a metric and the features which are not having high score are removed. Subset selection method will select the optimum features from the entire feature space.

Feature selection algorithms have been reviewed in Liu and Yu (Huan Liu & Lei Yu 2005). Many feature selection algorithms involve heuristic or random search strategies in order to reduce the computing time. Recently, many optimization algorithms have been used for feature selection, based on the level of importance, or based on the recognition percentage, they are producing. Optimization algorithms such as Genetic Algorithm (GA), Ant Colony Optimization (ACO) and Particle swarm Optimization (PSO) have been proposed for feature selection in speaker verification systems.

In our work, we have used GA (Genetic Algorithm) to find the best combination of features. The combinations of features are selected such that they must provide enough speaker specific information including vocal tract and source excitation characteristics. MFCC feature represents the vocal tract characteristic of the speaker whereas the wavelet coefficients of linear predictive residual signal represent the characteristics of excitation signal. Thus these two features are said to be complementary to each other. Therefore the feature combinations used in our research are: (1) MFCC with WPT-LPR (Wavelet Packet Transform of Linear Predictive Residual) (2) RASTA-PLP with WPT-LPR, and (3) RASTA-PLP with Glottal Flow Derivative (GFD) (4) MFCC with GFD (Pati & Mahadeva Prasanna 2012). The objective is to obtain a set of four complementary feature extractors. The main idea is to select the best combination of those features and as well as the set of features which are effective in producing the good performance using GA.
Conventional speaker verification system adopts the Mel Frequency Cepstral Coefficients (MFCC), Perceptual Linear Predictive coefficients (PLP), RASTA-PLP and Linear Predictive Cepstral Coefficients (LPCC) as the representative features. Each feature set has its unique characteristics. The extraction of these features has been explained in chapter 2. The combinations of these features are said to be a subset of features. The feature subset has various combinations. When different features are combined, they give a better performance of the speaker verification system than the individual features. Hence, GA can be used to select the best combination of feature sets. The performance of the system is evaluated with the TIMIT corpora.

3.2 GENETIC ALGORITHM

Genetic algorithms (GA) were first proposed by Holland (1975) and became widely used in various disciplines as a new means of complex systems optimization. In recent years, they have been successfully applied to the speech processing domain. GAs most attractive quality is certainly their aptitude to avoid local minima. However, our study relies on another quality which is the fact that GA is unsupervised optimization method.

So they can be used as an exploration tool, free to find the best solution without any constraint. State of the art speaker verification systems are based on a cepstral feature extraction front end (LFCC, MFCC, LPCC) followed by a GMM or an hybrid GMM/SVM classifier (Campbell et al 2006). Nowadays, an alternative approach is fusing of different systems. Fusion of systems may be based on the classifier or based on the features used.

Genetic algorithm is a random search algorithm (Whitley 1993) inspired by natural evolution. This algorithm uses set of solutions for the problem in the form of chromosomes. The chromosomes are generated by
encoding the solutions. The number of values assigned for individual chromosome depends on the number of components in the solution. After processing each chromosome is assigned with a fitness value and the best chromosome will be selected based on the fitness value. It implies that the best and better solutions to the problem are used to reproduce the other possible solutions and least fitted solutions are neglected from the generation of next population. This selection of best and better solutions based on the principle of “survival of the fittest”.

The execution of the Genetic Algorithm can be seen as a two stage process: It starts with the current population. Pairs of individuals are selected based on their fitness value to create an intermediate population. There exist different selection methods. The easiest one is to map the entire population onto a roulette wheel, where each individual is represented on the wheel by a space that is proportional to its fitness. The roulette wheel is rotated and an individual is chosen until the intermediate population is filled up. Then the genomes of each pair of the intermediate population perform a crossover and mutation. That means they are exchanging parts of their genetic information, i.e. bits of the number which represents the genome, and afterwards each bit is flipped with a small probability (e.g. 1%). The simplest way to perform a crossover is to randomly choose a crossover point, cut both chromosomes at this point, and swap the fragments between the two parents. Thus by recombining e.g. (11110000) and (00001111) with crossover point 4 one would get (11111111) and (00000000). All the children which are obtained by crossover and mutation are inserted into the next generation. Then the algorithm restarts using the next population as the current population. The process of evaluation, selection, recombination and mutation forms one generation in the execution of the Genetic Algorithm. This process is illustrated in Figure 3.1.
Major components of GAs include encoding schemes, fitness evaluations, parent selection, crossover operators, and mutation operators.

### 3.2.1 Gene Encoding

Parameter’s encoding plays a major role in genetic algorithm. The easiest and most common way to do this is by means of a binary code, but other coding is also possible. It reduces the over fitting effect by reducing the parameters dimension. Each feature vector can be coded as a ‘n’ bit binary number. Putting ‘p’ feature vectors, each of size ‘n’ bits together we get a population. Each individual feature vector of population is referred to as a chromosome. Each chromosome corresponds to a certain value of the evaluation function. This is expressed, by means of the fitness value, which is unique to each individual. The fitness function maps the outputs of the evaluation function for the different individuals in the population.
3.2.2  **Fitness Evaluation**

After the generation of a solution population, the fitness value of each member in the population is calculated. For a maximization problem, the fitness value $f_i$ of the $i$th member is usually the objective function evaluated at this member. Then the members are ranked according to their fitness value. The advantage of this is that the objective function does not need to be accurate, as long as it can provide the correct ranking information.

3.2.3  **Selection**

After ranking the members of initial population through fitness evaluation, the next population is generated from the current population through selection process. Selection operation determines which parents participate in producing offspring for the next generation, and it is based on *survival of the fittest* in natural selection. Usually, selection probability will be assigned to each member proportional to their fitness value. Then the members for mating are selected based on the selection probability equal to $f_i / \sum_{k=1}^{n} f_k$ where $n$ is the population size. The effect of this selection process is to retain the members with above average fitness values in the next generation and to replace the members with below average fitness values in the former one.

3.2.4  **Crossover**

In order to extract the good features of the previous generation chromosomes, Crossover is performed between them. Crossover is usually applied to the selected pairs of parents with a probability equal to a given crossover rate. The crossover operator exchanges substrings in two selected chromosomes and produces a pair of new chromosomes. The effect of crossover is similar to that of mating in the natural evolutionary process, in
which parents pass segments of their own chromosomes on to their children. Therefore, children may outperform their parents if they have good characteristics received from both the parents.

### 3.2.5 Mutation

Mutation is an operator that changes one or more genes of a chromosome in order to produce a population with new characteristics. If the population generated does not have enough mixing that can produce a satisfactory solution, mutation operation is done. Mutation is capable of spontaneously generating new chromosomes. Mutation is an important part of the genetic search helps to prevent the population from stagnating at any local optima.

Mutation is performed during evolution according to a user defined mutation probability. Usually the mutation probability is set as 0.01. This is accomplished by following different ways:

**Flip Bit** - A mutation operator that simply inverts the value of the selected gene (0 is changed as 1 and 1 is changed as 0). This mutation operator can only be used for binary genes.

**Boundary** - A mutation operator that replaces the value of the selected gene with either the upper or lower bound for that gene. This mutation operator can only be used for integer and float genes.

**Non-Uniform** – In this type of mutation, the probability that the amount of the mutation increases as the generation number increases. It maintains the population in the early stages of the evolution then allows fine tuning the solution in the later stages of evolution.
**Uniform** - A mutation operator that changes the value of the selected gene with a uniform random value chosen between the user-defined upper and lower bounds for that gene.

**Gaussian** - A mutation operator that adds a unit Gaussian distributed random value to the chosen gene.

### 3.3 FEATURE SELECTION USING GENETIC ALGORITHM

#### 3.3.1 Hybrid Feature Selection Using Genetic Algorithm

In our application, each individual of the population represents the feature combination and the particular feature extracted from a particular frame.

![Figure 3.2 Feature selection using genetic algorithm](image)

The components of chromosomes are the binary numbers which represent the feature combination and the selection of feature vector from a pool of input feature vectors. Figure 3.2 shows the proposed system for the feature selection using GA. The GA starts with the initial population which consists of set of random binary chromosomes. The first two binary digits represents the features which are to be combined (for example, 00-MFCC + WPTLPR;
01-MFCC+GPD; 10-RASTAPLP+WPTLPR; and 11-RASTAPLP+GPD). Half of the remaining bits represent the feature vectors of first kind and the remaining bits represent the feature vectors of second kind. The size of chromosome depends on the number of feature vectors in a feature set. i.e., each gene of the chromosome represents a particular feature vector. The binary value of the gene indicates whether the feature vector is included in the evaluation process or not. ‘1’ indicates the inclusion of feature vector and ‘0’ indicates that the feature vector is not included. For example, the chromosome “00101” indicates the inclusion of the feature vectors corresponding to the third and fifth frame of a particular feature extracted from the speech frames. The two different features are combined and it is being used for the system development. All experiments we made are based on a state of the art GMM speaker verification system. We used a system with 16 Gaussian mixtures, with diagonal covariance matrix. The optimal individuals are selected from the population by using an objective function, which measures how much it can be fit to the speaker model, and repeats the evolutionary process, which makes it get better. The processes involved in the feature selection using GA for GMM based speaker verification system are described in the Figure 3.3.

The objective function is log likelihood score for each individual feature vector. This in turn represents a particular combination of feature sets. Given a stream of data \( X_1, X_2, \ldots, X_T \), the score that correlates to the fit of the data with model can be computed as

\[
\text{Log Likelihood Score} = \sum_{i=1}^{T} \log \left( \sum_{k=1}^{M} p_k g_k(x_i) \right) \quad (3.1)
\]
where \( p_k \) is the probability of individual Gaussian mixture density.

\( g_k(x) \) is the Gaussian function.

After this process, the GA finds an optimum feature set based on the output of objective function. The best feature set has the maximum log likelihood score. The fitness function is a function of log likelihood ratios. Then, the stopping criteria are checked. If 90% of the feature vectors, having the same maximum log likelihood score, or the specified number of generations reached, then, GA reaches the optimum level. If not, crossover and mutation are carried out according to specified crossover and mutation
rate. Then, the process is repeated from the evaluation step with the new population until a termination condition is satisfied.

Now, from the best chromosome, the best combination of features and the feature vectors of those feature sets are selected. Here, the number of feature vectors selected for optimization is decided by a batch by batch procedure. That means, at a particular time, a set of feature vectors corresponding to a few seconds are considered for optimization (for example 100 feature vectors of MFCC/RASTAPLP). The whole set of training vectors are grouped into subsets. Optimization using GA is applied over all subset of feature vectors. For each subset, the feature vector corresponding to the best chromosome has been collected and the set of optimized feature vectors are stored as code vectors for the speaker.

During testing, the best combination of features determined during training for that speaker is used for feature extraction. Those features are extracted from the test speech frames and are compared with the feature code vectors assigned for that speaker. The feature vectors which are closest to the feature code vectors are only selected for testing. This method of training and testing of speaker model using best feature vectors and best combination of feature vectors leads to effective speaker model.

3.3.2 Feature Selection Using Genetic Algorithm with Score Fusion

In this proposed method, the individual feature sets are used by GA for feature selection. GA is used to determine the optimum WPTLPR features and RASTA-PLP features separately. Then, those features are used to develop a GMM model separately. A GMM-UBM for Speakers using WPTLPR features and a GMM-UBM for speakers using RASTA-PLP are developed. During training the speaker GMM models are obtained through MAP adaptation, whereas, during testing the optimal test feature vectors are applied
to the claimed speaker model. Then the scores from individual feature set based systems are combined with some weights. The fused score with confidence measure is used to make the decision. This method of feature selection using GA and score fusion is shown in Figure 3.4.

**Figure 3.4** Feature selection using GA and score fusion for speaker verification system
The log likelihood score is calculated separately for each feature set. Since each feature has its own characteristics, the score fusion with confidence measure (CM) improves the performance of the system.

First the speaker discrimination power of each feature set is determined from their log likelihoods using the equation

\[
D_i = \frac{\log P(s_i | \hat{\lambda}_{e,i}) - \log P(s_i | \hat{\lambda}_{u,i})}{\log P(s_i | \hat{\lambda}_{u,i})}, \quad i = 1, 2
\]

(3.2)

Here \(i = 1\) denotes the one feature set and \(i = 2\) denotes another feature set. If \(D_i > 0\) or \(D_i < 0\), the features are considered to have high confidence in the decision of accepting a genuine speaker or rejecting an impostor. If \(D_i\) is close to 0, the confidence is low and the decision tends to be uncertain. This discrimination power-based score fusion technique gives a much improved overall verification accuracy.

### 3.4 SUMMARY

Feature selection provides the improvement in the performance of the system and reduces the computation time. Among various optimization methods used for feature selection, this chapter describes the use of Genetic algorithm for feature selection. The selection of two different feature set and the selection of optimum feature vectors from the feature sets are performed by GA. The best combination of features selected by GA is GPD and RASTA-PLP.