Priyadarshani, P.G.N. & N.G.J. Dias,
Department of Statistics & Computer Science, University of Kelaniya
Paper: Equity

Genetic algorithm approach for sinhala speech recognition

Speech recognition is the ability to understand the spoken words and convert them into text. Nowadays there is a considerable tendency of developing ASR systems which are capable of tracking the human speech done in local specific languages and identifying them because the people prefer to use their native language. Even though there is a dire need of Sinhala speech recognition, it is still in the beginning. Here we have applied Genetic Algorithm (GA) for automatic recognition of isolated Sinhala words and Mel Frequency Cepstral Coefficients (MFCC) to model the speech signal.

GA is not considered as a mathematically guided algorithm. In fact, GA is a stochastic nonlinear process. Generally, GA involves a three operation selection of crossover and mutation that emulate the natural genetic behavior. The purpose of selection is to determine the genes to retain or delete for each generation based on their degree of fitness. Even though there are several types of selection methods, we have used elitist selection as we observed that it allows to retain a number of best individuals for the next generation and improve the recognition capability. If the individuals are not selected to reproduce they may be lost. But the fittest individual survives. Crossover (reproduction) is a process to exchange chromosomes to create the next generation. Rather than two-point and uniform crossover, in this work we have used one point crossover with probability 0.80 to prevent unnecessary crossover. A mutation is a change of a gene found in a locus randomly determined. The altered gene may cause an increase or a weakening of the recognition. Mutation probability is usually very low. Each offspring is subjected to mutation with probability 0.01.

The reference dictionary (learning corpora) is the population managed by our genetic algorithm. Initially we selected ten Sinhala words as the vocabulary with 24 repetitions for each word from three speakers. Therefore, the dictionary is made up of 240 individuals. This population is divided into 10 sub-populations (the number of words), the choice of the initial population is random for each word to be recognized. An initial population is made up of all occurrences of a word, i.e., 24 individuals.

To evaluate the performance, we carried out two types of tests. We used 6 repetitions of each word made by three speakers who participated in the learning process and 10 repetitions of each word generated by a completely new speaker. First test proved that our GA is capable of handling multiple speakers. And the second test proved that our GA is independent of the speaker. Further, word recognition of registered speakers is dominant compared to a relatively unregistered speaker. However, results indicated a satisfactory precision even for speaker independent cases.

Keywords: **MFCC, Genetic algorithm, selection, crossover, mutation, threshold**