Original Article

Genetic algorithm used in Speech Recognition methods

A V Wadagale[1*], V A Jadhav[2]

[1]Ph.D. Research Student, [2]Associate Professor, Department of Mathematics, Statistics and Computer Sciences, NES Science College, Maharashtra, INDIA.
Email: atulvw@gmail.com

Abstract

Speech recognition by use of GA, its background and introduction., There are two classic approaches for development of Speech recognition as: 1. Dynamic Time Warping (DTW) 2. Hidden Markov Model (HMM). Discussion on these two with suitable example is explain in the given paper.
Key words: Genetic algorithm, Speech Recognition.

***Address for Correspondence:**
Dr. A V Wadagale, Ph.D. Research Student, Department of Mathematics, Statistics and Computer Sciences, NES Science College, Maharashtra, INDIA.
Email: atulvw@gmail.com
Received Date: 15/12/2015 Revised Date: 13/01/2016 Accepted Date: 22/02/2016

Access this article online	
Quick Response Code:	Website: www.statperson.com
	DOI: 24 February 2016

INTRODUCTION

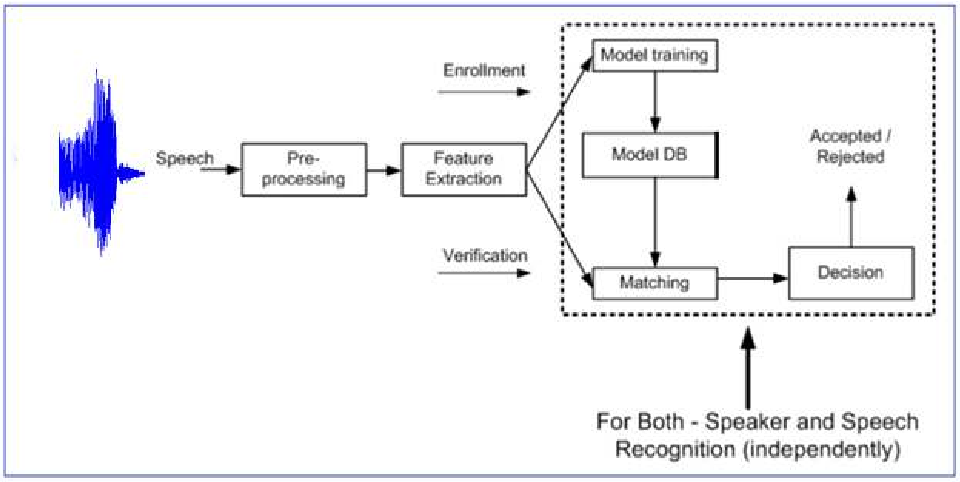
Speech; It defines as the expression of or the ability to express our thoughts and feelings by articulate sounds. In today's world, we all communicate with each other via this natural phenomenon called speech. That something we spoken or exchange of words and conversation usually interacting with each other and replying them by identifying the word they spoke is call recognition. In speech recognition, the training process plays vital role. The training process sometimes also called as learning process. When a good training model come into possession for a speech pattern then this is further improved the quality of speed of recognition and also become better the quality of the overall performance of speech recognizing. Signals of speech are composition of sequence of sound. The transitions between these sounds are like symbolic presentation of given information. The arrangement in sequence of these sounds is organized by rule of language. Linguistics is the study of all if this i.e. the study of these rules and implementation of it in human communication. Phonetics is the study of the sounds of the speech. Representation of the speech can be done by given massage content. The acoustic waveform is alternative way for characterizing of speech in terms of signal carrying its massage information. For human and his environment, speech plays as important role for communication and that's why Automatic System Recognition (ASR) is essential for human for all time. There are different types of speech with isolated words, continuous words and connected words.

In literature, there are two classic approaches for development of Speech recognition as follows:

1. Dynamic Time Warping (DTW)
2. Hidden Markov Model (HMM).

In DTW, GA is applied to solve discrete, constrained problems and nonlinear problem in HMM, using parallel GA, a unique approach is obtain a HMM model. As properties of GA belonging naturally, without extra computational cost, the associated non-trival K-best paths of DTW can be identified. An abstract idea of parallel GA used in HMM is recognized as being a sound way for the process of discovering the best and if not most favorable template reference model.

Because of the research in advance technology that have been made in the areas of speech processing and pattern recognition, the rare possible dream of speaking to computers only have come true. Given two technologies classify the label of each particular utterance and allowed to model speech utterances in a correct manner. Main task of speech recognition is essential in capturing a speech signal from an input device like microphone and then it is identifying the label of this signal. This task given is simple in manner because of the different time varying properties contain in speech signals. For example, if a person speaks the one same word for many times, his utterance will come with different characteristics. When in Marathi word Ek i.e. 'One' was spoken five times, it can be seen from Fig. 5.1 The speech utterances did not have the same output across the range of whole utterance. Thus, in the course of developing a speech recognition system, the crux of matter would seem to be to perform a robust and accurate matching in the recognition process. This is by no means an easy task, as speech utterance in general is fuzzy, varies in time and is sometimes unpredictable.



Years of hard work by the researchers in the field of Speech recognition, obviously, much speech recognition system software and hardware have been successfully launched. For Examples, the automation of directory services by NYNEX and Northern Telecom, the automation of operator services by AT&T and Northern Telecom, the stock quotation services by Bell Northern Research. Now we focus on the application of the Genetic Algorithm to speech recognition systems instead of looking back to history and review of Speech recognition in details. Also we will deal with the improvement the performances of the speech recognition system with special area of matching of speech pattern.

In speech recognition, speech utterances shall convert to digital presentations for starting computer processing. Therefore, through transducer such as microphone speech signals i.e. sound waves will pass from it and be digitization of it into a set of discrete data points. This conversion must be accurate enough to represent the original signal and ensure that the integrity of the speech signals is preserved. As per Shannon's sampling theorem, the originality of the signal will not be distorted if sample signals at the Nyquist rate i.e. it is equal to two times the bandwidth of the original signal of process. It is well known that human speech is lies below 4KHz, that's why, a sample rate of 8KHz is very much sufficient to reproduce the original signal without loss of originality. For conversion of analog speech signal into digital form, an analog to digital converter with 12bits resolution and 16,000hz of sampling frequency can be used.

Figure 5.1: Block diagram for Speech Recognition System

As the raw speech signals are digitized, discrete points set is then converted to feature vectors so that the speech signal can be best represented with a minimum number of points. In fact, the purpose of feature analysis is also to distill the necessary information from the raw speech signals. The feature vectors can be obtained either in time domain. In time domain, the feature vectors provide useful representation such as energy, pitch, zero crossings. However, these measurements do not reveal the entire picture of speech signals completely. Sometimes, given measurements can only used to check the voiced and non-voiced parts of speech signals and also noise for the end-point detection of difference between signals.

Over a short period of time ranging between l0ms to 30ms, also spectral analysis methods can applied to speech signals in order to extract the short time spectral covers characteristics of the signal. There are many techniques which can be used for spectral analysis such as Fast Fourier transform (FFT), linear prediction and Spectral analysis, Discrete Fourier transform (DFT), filter bank. Among these techniques, Atal et al. has pointed out that Spectral analysis has been found to be one of the most effective feature vectors that can provide the best results. One example is the HTK recognizer developed in Cambridge University in which researchers used the 12 mel frequency cepstral coefficients (MFCC) and the signal energy to form the basic 13-element acoustic feature vector. In addition, they also demonstrated empirically that the MFCC can improve the system performance of most speech recognition systems.

Once the process of extracting feature vector has completed, two other modes of operation have to be performed in speech recognition systems, namely (1) training and (2) recognition. These operations ensure that the speech utterance are labeled and stored in a database for the purpose of recognition.

As each utterance has its own distinct characteristics in terms of energy or spectral features, it is necessary to provide a way to find a single or multiple utterances to be best represented for all the members of the same class. This is known as a labeling process and in general it is called as Training or learning in a speech recognition system. To continue on from our previous example for the recognition of the Chinese word 'one', this particular utterance has to be trained and then stored in the database. In this case, the same word 'one' with four similar but varying sets of feature vectors can be used for training.

REFERENCES

1. Antonisse, J. 1989. A new interpretation of schema notation that overturns the binary encoding constraint. In J. D. Schaffer, ed., Proceedings of the Third International Conference on Genetic Algorithms. Morgan Kaufmann.
2. Axelrod, R. 1984. The Evolution of Cooperation. Basic Books.
3. Axelrod, R. 1987. The evolution of strategies in the iterated Prisoner's Dilemma. In L. D. Davis, ed., Genetic Algorithms and Simulated Annealing. Morgan Kaufmann
4. Axelrod, R., and Dion,D. 1988. The further evolution of cooperation. Science 242, no. 4884: 1385–1390.
5. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walters. The Shape and Structure of Proteins. Garland Science, New York and London, fourth edition, 2002.
6. Bäck, T. 1994. Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In Proc. 1st IEEE Conf. on Evolutionary Computation. Piscataway, NJ: IEEE Press, pp. 57–62.
7. Davis, L. 1991. Hybridization and numerical representation. In L. Davis (ed.), The Handbook of Genetic Algorithms. New York:Van Nostrand Reinhold, pp. 61–71.
8. De Jong, K. A. 1975. An Analysis of the Behavior of a Class of Genetic Adaptive Systems. Ph.D. thesis, University of Michigan, Ann Arbor.
9. Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi. 1983. Optimization by simulated annealing. Science 220:671–680.
10. Linden, D. S. (1997). Automated Design and Optimization of Wire Antennas using Genetic Algorithms. PhD thesis, MIT
11. Pierre, D. A. 1992. Optimization. In McGraw-Hill Encyclopedia of Science and Technology 12. New York: McGraw-Hill, pp. 476–482.
12. Powell, M. J.D. 1964. An efficient way for finding the minimum of a function of several variables without calculating derivatives. Comput. J. 7:155–162.

Source of Support: None Declared
Conflict of Interest: None Declared