SPEECH RECOGNITION AND HIDDEN MARKOV MODEL

Shahin Samimi*

Fariba Rastegar*

Abstract: *The discussion of automatic identification of speech can be studied due to two aspects of producing speech and understanding and receiving speech. The hidden model of Markov (HMM) is an effort for statistical modeling of the speech producing system. So it is belonged to the first group of speech identification methods. It is important that how HMM model is compatible with the accidental nature of characteristic vector quantity.*

Keywords: *Hidden Morkov Model, Speech Recognition, audio components*

*Department of Computer Engineering, Behbahan Branch, Islamic Azad University, Behbahan, Iran

1– INTRODUCTION

In addition to speech recognition capability, computers should be able to express and talk in order to establish two-way communication between computer and human. This provides the machine the ability to read electronic texts and letters or to express words and answers. This system alone could, for example, be used in the cars or at home for reading newspapers, or be beside the speech recognition system or integrated with it. For example, in the automatically telephonic secretary system, the computer gives him some guidance or an appropriate response after recognizing the client's speech. One of the issues in signal processing that has gained much attention is signal modeling. There are several options for modeling a signal and its characteristics. From one perspective; we can divide signal models into two categories: specific or definite models and statistical models. Statistical models try to create a model by the use of statistical properties of the signal. Gaussian models, Markov chain and Hidden Markov models are among these approaches. The basic assumption in the statistical models is that the properties of a signal can be modeled as a parametric random process.

2-THE SPEECH RECOGNITION STAGES

2-1- user input

At this stage, the user expresses his request in some word or phrases. Speech recognition system records the user's voice into an analog audio signal.

2-2- Numerical simulation (digitization)

At this stage, the analog audio signal is converted into a digital signal.

2-3- Analysis of audio components

The language consists of a collection of different sounds each of which is called a phoneme. A syllable is made of a combination of phonemes and a combination of syllables makes the words. Each phoneme can be identified as a specific pattern in the spectrogram.

Phoneme detection requires intense concentration on sound energy which is called Formant that, in every frequency, has increase and gradual decrease characteristics and it is one of the most noticeable features of the human voice.

Although phonemes are not recognized in a sound wave, the acoustic wave can be decomposed into its constituent frequencies and displayed in a spectrogram. The vertical axis in the spectrogram shows the frequencies above 8000 Hz and the horizontal axis indicates the passing of the time.

2-4- Statistical Modeling

After the above steps, the system starts to be involved in matching the sounds and voices with the sounds and voices defined in itself. A dictionary is used to show how to pronounce a word and Speech recognition machine uses this dictionary.

2-5- Making adjustment

Modeling language technology is applied to increase the accuracy. By a list of rules, recognized sounds can predict their following noises. In this system a list of words or phrases matched with the mentioned sentences is returned along with confidence coefficient.

Using Hidden Markov model-HMM is one of the most common ways of implementing. Hidden Markov model is based on the Mathematical models of digital signal processing and describes a complex system based on a finite set of states. These states express the circumstances of going from one state to the next.

3- HIDDEN MARKOV MODEL (HMM)

Here, the Markov model was introduced, in which each state corresponds to an observable event [1].

In this section we extend the above definition to the case where the observations of the functions are probabilities of the states. The resulting model is a stochastic model with an underlying stochastic process that is hidden and is visible only by a set of stochastic processes that produce the sequence of observations.

-The number of possible states:

The number of states has an important role in the success of the model and each state has a corresponding event in the Hidden Markov model.

-The number of observations in each state:

The number of observations equals to the outputs of a modeled system.

-Models or states of Model N:

It's the number of observation symbols in the alphabet, M. If the observations are discrete, then M will be unlimited.

$$\Lambda = \{a_{ij}\}$$

-Transition matrix of state *A*=[*a_{ij}*] .A set of probabilities of the transition between states:

$$a_{ij} = p\{q_{t+1} + 1 = j \mid q_t = i\}, \qquad 1 \leq i,j \leq N$$

where q_t represents the current state. Transition probabilities must satisfy the natural limitations of a random probability distribution. These constraints are as follows:

$$a_{ij} \geq 0, \qquad 1 \leq i,j \leq N$$

$$\sum_{j=1}^{N} a_{ij} = 1, \qquad 1 \leq i \leq N$$

In all states of the ergodic model, the value of a_{ij} is greater than zero for all *j* s and *i* s. In case there is no coupling (relations or connections) between states, we have a_{ij} = 0.

-Distribution of the probability of the observations:

$$B = \{b_j(k)\}$$

$$b_j(k) = p\{o_t = v_k | q_t = j\}, \ 1 \leq j \leq N, \ 1 \leq k \leq M$$

It's a probability distribution for each of the states where V_k represents k^{th}, the observed symbol in the alphabet and o_t indicates the vector of the current input parameters. For the values of the probability of the states, the conditions in the probability theory should also be respected [2].

$$b_j(k) \geq 0, \ 1 \leq j \leq N, \ 1 \leq k \leq M$$

$$\sum_{k=1}^{M} b_j(k) = 1\,, \ 1 \leq k \leq M$$

If the observations are continuous, a continuous probability density function must be used rather than discrete probabilities.

Usually the probability density is estimated by a weighted sum of M and normal distribution *N*.

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} N(\mu_{jm}, \sum jm\,, o_t)$$

In which C_{jm}, μ_{jm} and $\sum_{jm}$ are respectively the weighting coefficient, the mean vector and the covariance matrix. In the above equation, the values of C_{jm} must satisfy the following conditions:

$$c_{jm} \geq 0,\ \ 1 \leq j \leq N,\ \ 1 \leq m \leq M$$

$$\sum_{m=1}^{M} c_{jm} = 1,\ \ 1 \leq j \leq N$$

The probability distribution of the initial state $\pi = \{\pi_i\}$ in which

$$\pi_i = p\{q_1 = i\},\ \ 1 \leq i \leq N$$

So, the Hidden Markov model with a discrete probability distribution can be recognized by the following triad:

$$\lambda = (\Lambda, B, \pi)$$

Also, the hidden Markov model with a continuous probability distribution is shown as follows:

$$\lambda = (\Lambda, c_{jm}, \mu_{jm}, \Sigma_{jm}, \pi)$$

4- TYPES OF HIDDEN MARKOV MODELS AND CONTINUOUS HMM:

Structurally and in terms of topology, Hidden Markov Model is of different types:

As mentioned, in the Ergodic model $a_{ij} > 0$ is for all *i* s and *j* s. The model structure is like a perfect speech in which the vertices are connected recursively (they have returning connections). However, due to the complexity of the process, different and special structures are needed for different applications [3]. Among these structures that are widely used in the applications of the speech recognition based on the phoneme and of the speaker recognition, is the Left-right model 1 or the Bkys model 2. This model, whose structure can be seen in Figure 2, contains left-to-right connections and is used for modeling signals whose properties change during the time. There is only one input state in the left-to-right model that is the first state, and so:

$$\pi_t = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

Ergodic and left-to-right models are based HMM models and have the most applications in speech processing, although it is possible to create more flexible models by connecting several models or changing the structure of its connections.

Figure 1-C shows a model of a typical left-to-right shunt, which includes two models of left-to-right

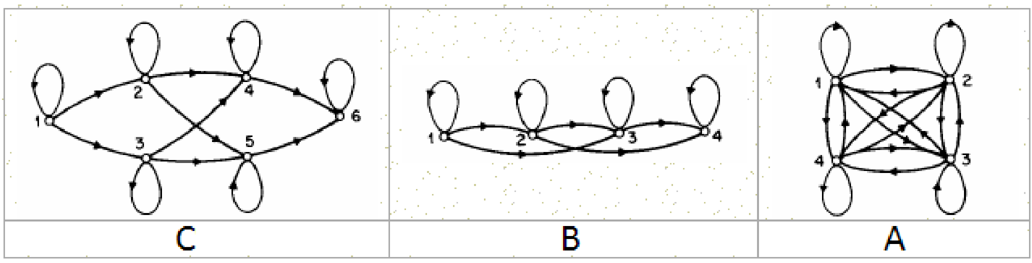

Figure 1: 3 structure for HMM model

a) Ergodic HMM model b) left to right model c) right to left parallel model

In the previous sections, we examined the HMM models for a set of discrete observations [4]. Although it is possible to convert all continuous processes into the processes associated with a sequence of discrete observations by quantifying, it may cause model dropping (model is impaired). In the continuous HMM model, the Probability of being in a state for the observations is shown by the probability density functions. In these conditions, for any input mode, the probability of observing $b_t(O)$ is shown as a distribution including mixture *M* for each *i* state and *O* input:

$$b_t(O) = \sum_{m=1}^{M} c_{tm} \Re(O, \mu_{tm}, U_{tm})$$

In which C_{tm} is the mixing factor (coefficient of mixture) *m* and □ can be any density function. Gaussian function is usually used for this purpose.

The mixing coefficients (mixing factors) must have the following restrictions (limitations):

$$\sum_{m=1}^{M} c_{tm} = 1, \quad 1 \leq i \leq N$$

$$c_{tm} \geq 0 \quad 1 \leq i \leq N \quad 1 \leq m \leq M$$

5- CONCLUSION

During recent years, HMM method as the most successful method in identification of speech is used. The main reason is that HMM. Model is able to define the characteristics of speech signal in the form of understandable mathematic HMM model can be designed in the way that receive every one of these different inputs.

REFERENCES

[1] Rabiner, L. and Wilpon, J. and Soong, F. (1988), “High Performance Connected Digit Recognition using Hidden Markov Models”, IEEE Transaction of Acoustic, Speech, and Signal Processing, Vol. 37, No. 8, pp. 1214-1225.

[2] Flahert, M.J. and Sidney, T. (1994), “Real Time implementation of HMM speech recognition for telecommunication applications”, in proceedings of IEEE International Conference on Acustics, Speech, and Signal Processing, (ICASSP), Vol. 6, pp. 145-148.

[3] Anusuya and Katti (2009), “Speech Recognition by Machine: A Review”, International Journal of Computer Science and Information Security, Vol. 6, No. 3, pp.181-205.

[4] Gaikwad, Gawali and Yannawar(2010), “A Review on Speech Recognition Technique”, International Journal of Computer Applications, Vol. 10, No.3, pp. 16-24.