Speech Recognition with Hidden Markov Model: A Review

Shivam Sharma

Abstract: The concept of Recognition one phase of Speech Recognition Process using Hidden Markov Model has been discussed in this paper. Preprocessing, Feature Extraction and Recognition three steps and Hidden Markov Model (used in recognition phase) are used to complete Automatic Speech Recognition System. Today's life human is able to interact with computer hardware and related machines in their own language. Research followers are trying to develop a perfect ASR system because we have all these advancements in ASR and research in digital signal processing but computer machines are unable to match the performance of their human utterances in terms of accuracy of matching and speed of response. In case of speech recognition the research followers are mainly using three different approaches namely Acoustic phonetic approach, Knowledge based approach and Pattern recognition approach. This paper's study is based on pattern recognition approach and the third phase of speech recognition process 'Recognition' and Hidden Markov Model is studied in detail.

Keywords: Automatic Speech Recognition (ASR), HMM model, human machine interface.

1. Recognition

Recognizers is the third phase of speech recognition process deal with speech variability and account for learning the relationship between specific utterances and the corresponding word or words [1]. There has been steady progress in the field of speech recognition over the recent yeas with two trends [2]. First is academic approach that is achieved by improving technology mainly in the stochastic modeling, search and neural networks. Second is the pragmatic, include the technology, which provides the simple low-level interaction with machine, replacing with buttons and switches. A second approach is useful now, while the former mainly make promises for the future. In the pragmatic system emphasis has been on accuracy, robustness and on the computational efficiency permitting real time performance with affordable hardware. Broadly speaking, there are three approaches to speech recognition [3] [4].

(a) Acoustic-phonetic approach: Acoustic-phonetic approach assumes that the phonetic units are broadly characterized by a set of features such as format frequency, voiced/unvoiced and pitch. These features are extracted from the speech signal and are used to segment and level the speech.

(b) Knowledge based approach: Knowledge based approach attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analyzing and finally making a decision on the measured acoustic features. Expert system is used widely in this approach.

(c) Pattern recognition approach: Pattern recognition approach requires no explicit knowledge of speech. This approach has two steps – namely, training of speech patterns based on some generic spectral parameter set and recognition of patterns via pattern comparison. The popular pattern recognition techniques include template matching, Hidden Markov Model [5].

2. Hidden Markov Models (HMM)

HMM is doubly stochastic process with an underlying stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce sequence of observed symbols. The basic theory behind the Hidden Markov Models (HMM) dates back to the late 1900s when Russian statistician Andrej Markov first presented Markov chains. Baum and his colleagues introduced the Hidden Markov Model as an extension to the first-order stochastic Markov process and developed an efficient method for optimizing the HMM parameter estimation in the late 1960s and early 1970s. Baker at Carnegie Mellon University and Jelinek at IBM provided the first HMM implementations to speech processing applications in the 1970s [6]. Proper credit should also be given to Jank ferguson at the Institute for defense Analysis for explaining the theoretical aspects of three central problems associated with HMMs, which will be further discussed in the following sections [7]. The technique of

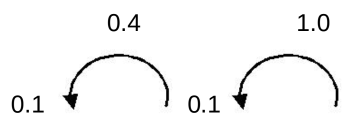
HMM has been broadly accepted in today's modern state-or-the art ASR systems mainly for two reasons: its capability to model the non-linear dependencies of each speech unit on the adjacent units and a powerful set of analytical approaches provided for estimating model parameters [8] [9].

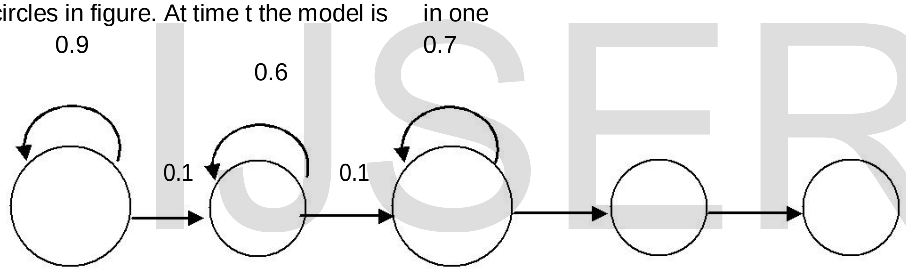
3. DEFINITION

The Hidden Markov Model (HMM) is a variant of a finite state machine having a set of hidden states Q, an output alphabet (observations) O, transition probabilities A, output (emission) probabilities B, and initial state probabilities II. The current state is not observable. Instead, each state produces an output with a certain probability (B). Usually the states Q, and outputs O, are understood, so an HMM is said to be a triple (A, B, II).

4. DESCRIPTION OF HMM

For the description figure 1 shows an example of Hidden Markov Model, The model consists of a number of states, shown as the circles in figure. At time t the model is in one of these states and outputs an observation (A, B, C or D) [10] [11] At time t+1 the model moves to another state or stays in the same state and emits another observation. The transition between states is probabilistic and is based on the transition probabilities between states which are given in state j at time t+1. Notice that in this case A is upper triangular. While in a general HMM transitions may occur from may state to any other state, for speech recognition applications transitions only occur from left to right i.e. the process cannot go backwards in time, effectively modeling the temporal ordering of speech sounds. Since at each time step there must always be a transition from a state to a state each row of A must sum to a probability of 1. The output symbol at each time step is selected from a finite dictionary. This process is again probabilistic and is governed by the output probability matrix B where Bjk is the probability of being in state j and outputting symbol k. Again since there must always be an output symbol at time t, the rows of B sum to 1 [12]. Finally, the entry probability vector π, is used to described the probability of starting in described by the parameter set λ = [π, A, B]





$$A = \left[a_{ij}\right] = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 1.0 \end{bmatrix}$$

$$B = \left[b_{jk}\right] = \begin{bmatrix} 0.1 & 0.8 & 0.1 & 0.0 \\ 0.0 & 0.2 & 0.8 & 0.0 \\ 0.7 & 0.0 & 0.0 & 0.3 \\ 0.2 & 0.0 & 0.0 & 0.8 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$\pi = \begin{bmatrix} 0.4 & 0.6 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Figure 1: A Five State Left-Right, Discrete HMM for Four Output Symbols.

A HMM is characterized by the following:

□N, the number of states in the model. The individual states are denoted as S={S1,S2,….,Sn} and the system state at time t as q1.

□M, the number of distinct observation symbols per state, i.e. the discrete alphabet size. The individual symbols are denoted as V = {v1,v2,….,vm}

6

□ The transition probability distribution A={aij} where, each aij is the transition probability from state Si to state Sj. Clearly, aij ≥ 0 and □aij □1,□i
k

□ The observation symbol probability distribution B = bjk where, each bjk is the observation symbol probability for symbol vk, when the system is in the stae Sj. Cleary, bij ≥ 0, □j k and □b jk □1,□j
k

□The initial state distribution π={ π } where, π = P[q1 = S1], 1 ≤ j ≤ N. HMM model can be specified as λ = (A,B, π,M,N,V). In this thesis, HMM is represented as λ = (A, B, π) and assume M, N and V to be implicit.

5. USE OF HMM IN SPEECH RECOGNITION

HMM can be used to model a unit of speech whether it is a phoneme, or a word, or a sentence. LPC analysis followed by the vector quantization of the unit of speech, gives a sequence of symbols (VQ indices). HMM is one of the ways to capture the structure in this sequence of symbols. In order to use HMMs in speech recognition, one should have some means to achieve the following:

Evaluation: Given the observation sequence O = user

(o1, o2, …,oT) and a HMM λ = (A,B, π) to choose a corresponding state sequence Q = q1, q2,…,qT which optimal in some meaningful sense, given the HMM.

Training: To adjust the HMM parameters λ = (A,B, π) to maximize P(O | λ).

The following are some of the assumptions in the Hidden Markov Modeling for speech.

Successive observations (frames of speech) are

independent and therefore the probability of sequence of
T
….oT) = □P(oi) .
i□1

Markov assumption: The probability of being in a state at time t, depends only on the state at time t-1.

The problems associated with HMM are explained as follows:

(a) Evaluation: Evaluation is to find probability of generation of a given observation sequence by a given model. The recognition result will be the speech unit corresponding to the model that best matches among the

that is $\delta_t(i)$ is the best score along single path at time *t*, which accounts for the *t* observations and ends in state *i*. by induction,

$$\delta_{t+1}(j) = [\max_{i} \delta_t(i) a_{ij}]\, b_j(o_{t+1})$$

(c) Training (Learning): Learning is to adjust the model parameters (A, B, π) to maximize the probability of the observation sequence given the model. It is the most difficult task of the Hidden Markov Modeling, as there is no known analytical method to solve for the parameters in a maximum likelihood model. Instead, an iterative procedure should be used. Baum-Welch algorithm is the extensively used iterative procedure for choosing the model parameters. In this method, start with some initial estimates of the model parameters and modify the model parameters to maximize the training observation sequence in an iterative manner till the model parameters reach a critical value.

6. CONCLUSION

The conclusion of this study of recognition and hidden markov model has been carried out to develop a voice based machine interface system. In various applications we

observation P = (o1, o2, …,oT) can be written as a product of probabilities of individual observations, i.e. O = (o1, o2,
can use this user machine system and can take advantages as real interface, these application can be related with disable persons those are unable to operate computer through keyboard and mouse, these type of persons can use computer with the use of Automatic Speech Recognition system, with this system user can operate computer with their own voice commands (in case of speaker dependent and trained with its own voice samples). Second application for those computer users which are not comfortable with English language and feel good to work with their native language i.e. English, Punjabi, Hindi.

different competing models. Now to find *P(O | λ)*, the probability of observation sequence $O = (o_1, o_2, \ldots, o_T)$ given the model λ i.e. *P(O | λ)*.

(b) Decoding: Decoding is to find the single best state sequence, $Q = (q_1, q_2, \ldots, q_T)$, for the given observation sequence $O = (o_1, o_2, \ldots, o_T)$. Consider $\delta_t(i)$ defined as

$$\delta_t(i) = \max_{(q_1, q_2 \ldots q_{T-1})} P[q_1, q_2 \ldots q_t = i, o_1, o_2 \ldots o_t \mid \lambda]$$

7. REFERENCE:

1. 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.

2. AbdulKadir K, (2010)," *Recognition of Human Speech using q-Bernstein Polynominals",* International Journal of Computer Application, Vol. 2 – No. 5, pp. 22-28.

3. Reddy, R. (1976), "*Speech Recognition by Machine: A Review*", in proceedings of IEEE transaction , Vol. 64, No. 4, pp. 501-531.

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

5. Atal, Bishnu S. and Rabiner, Lawrence R. (1976), "*A Pattern Recognition Approach to Voiced- Unvoiced Classificaton with Application to Speech Recognition*", in proceedings of the IEEE International Conference on

Acoustic, Speech and Signal Processing (ICASSP'76), Pennsylvania, Vol. 24, No. 3, pp.201-212.

6. Rabiner, L. and Juang, B.H. (1986), "*An Introduction to Hidden Markov Models*", IEEE ASSP Magazine, Vol. 3, No.1, Part 1, pp. 4-16.

7. Rabiner, L. (1989), "*A Tutorial on Hidden Markov Models and selected Application in Speech Recognition",* in proceedings of IEEE, Vol. 77, No. 2, pp. 257-286.

8. Picone, J. (1990), "*Continues Speech Recognition using Hidden Markov Models*", IEEE ASSP Magazine, Vol. 7, Issue 3, pp. 26-41.

9. 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.

10. 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.

11. Rabiner, L. and Levinson, S. (1989), "*HMM Clustering for Connected Word Recognition*", in proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP), Glasgow, UK, Vol. 1, pp. 405-408.

12. Rabiner, L. and Levison, S. (1985), "*A Speaker-independent, Syntax-Directed, Connected Word Recognition System based on Hidden Markov Model and level building*", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 33, Issue 3, pp. 561-573.

8. ACKNOWLEDGEMENT:

I owe special debt of gratitude to Mr. Dhiraj Pandey, Department of Computer Science & Engineering, JSS Academy of Technical Education, Noida for her constant support and guidance throughout the course of work. Her sincerity, thoroughness and perseverance have been a constant source of inspiration for me. It is only her cognizant efforts that my endeavours have seen light of the day. I also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the study. Last but not the least, I acknowledge my friends for their contribution in the completion of the research work.

Shivam Sharma has completed his B.Tech, Computer Science and Engineering from JSS Academy of Technology, Noida PH-08010206467.
E-mail: sharmashivam2806@gmail.com