IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-4, Special Issue-7, Dec 2016 ISSN: 2347-2693

Survey On Speech Recognition Using Hidden MARKOV Model

Adarsh Pradhan[1], Nabanita Paul[2], Kakali Das[3], Rupam Jyoti Bordoloi[4], Dikhya Baruah[5]
Department of Computer Science & Engineering, Girijananda Chowdhury Institute of Management and Technology (GIMT), Guwahati, India [1,2,3,4,5]

Available online at: www.ijcseonline.org

***Abstract*: Modeling a signal model for recognizing speech is a tough job. Here we will find out how Hidden Markov Model (HMM) is used in modeling speech recognition application. The steps: Preprocessing, Feature Extraction and Recognition and Hidden Markov Model which used in recognition phase, are used to design a complete Automatic Speech Recognition System (ASR). Researchers are trying to model a perfect ASR system but computer machines are not able to match up to our expectations due to lack of accuracy of matching and inefficient speed of response. In speech recognition three different approaches which are broadly used are Acoustic phonetic approach, knowledge based approach and Pattern recognition approach. This study is based on pattern recognition approach using Hidden Markov Model which is studied in detail as HMMs are found to yield best performance among all the available techniques.**

***Keywords*: MARKOV Model, HMM,**

I. INTRODUCTION

Signals are basically discrete or continuous in nature. The source of the signal can be stationary or non-stationary. Signals can be pure or can have errors from other signal sources or due to transmission distortions [2, 4]. Signals can be characterized in terms of signal model which helps us with theoretical description about the signal processing model and to understand immensely about signal source without the availability of the signal source.

Definition:
Signal model can be classified in to two types [2, 4]:
• Deterministic Model: It makes full use of some properties of signals (for example, Amplitude of wave).
• Statistical Model: It takes into consideration the mathematical properties of a signal. Example of this type of model is Gaussian Model, Poisson Model, Markov Model and Hidden Markov model.

Speech signals are slowly time varying signals because, when observed over a small interval of time (between 5 and 100 ms), its characteristics are short-time stationary. But this is not the scenario when we observe a speech signal for a longer time perspective (>0.5 s). For such cases the signals characteristics are non-stationary.

Speech Recognition:
Speech recognition is a process of converting speech signal to a sequence of words which incorporates knowledge and research in linguistics, computer science and electrical engineering fields.

II. HIDDEN MARKOV MODELS (HMM) [1, 2, 3, 6]

HMM is statistical Markov Model with an underlying stochastic process which cannot be observed directly, but the sequence of observed symbols produced by this set of underlying stochastic processes can be observed. The approach of HMM in speech recognition has been broadly accepted in today's modern state-of-the art ASR systems.

The HMM (M) is a type of a finite state machine having a set of states N that are hidden, an output K, transition probabilities A, output probabilities B, and initial state probabilities π[1, 3]. The current state cannot be observed. Instead, each state produces an output with a certain probability (B) which is observable.

$$M=\{A,B,\pi\}$$

III. MAIN ISSUES OF USING HIDDEN MARKOV MODELS [2,3]

- **Evaluation problem:** If we are given the HMM, $M=(A, B, \pi)$ and the observation sequence $O=O_1,O_2,O_3,\ldots,O_K$, them what is the probability that model M has generated sequence O .
- **Decoding problem:** Given the HMM $M= (A, B,\pi)$ and the observation sequence $O=O_1,O_2,O_3,\ldots,O_K$, calculate the most likely sequence of hidden states S_i that produced this observation sequence O.
- **Learning problem:** Given some training observation sequences $O = O_1, O_2, O_3,... O_K$ and numbers of hidden and visible states determine

HMM parameters M=(A, B, π) that best fit training data.

Where $O=O_1,O_2,O_3,\ldots..,O_K$ denotes a sequence of observations.

IV SOLUTION TO THE ISSUES OF USING HMM

The Viterbi Algorithm [8]:
The Viterbi algorithm is a dynamic programming algorithm to find the most likely sequence of hidden states known as Viterbi path which results in a sequence of observed events, mainly in the context of Markov Information sources and hidden Markov models.

The terms Viterbi path and Viterbi algorithm are also applied to related dynamic programming algorithms which discover the single most likely explanation for an observation. For instance, in statistical parsing a dynamic programming algorithm can be used to discover the single most likely context-free derivation (parse) of a string, which is sometimes called the *Viterbi parse*. The Viterbi algorithm was proposed by Andrew Viterbi in 1967 as a decoding algorithm for convolutional codes over noisy digital communication links. The algorithm has found universal application in decoding the convolutional codes used in both CDMA and GSM digital cellular, dial-up modems, satellite, deep-space communications, and 802.11 wireless LANs. It is now also commonly used in speech recognition, speech synthesis, keyword spotting, computational linguistics, and bioinformatics. For example, in speech-to-text (speech recognition), the acoustic signal is treated as the observed sequence of events, and a string of text is considered to be the "hidden cause" of the acoustic signal. The Viterbi algorithm finds the most likely string of text given the acoustic signal.

Algorithm [9]
Suppose we are given a Hidden Markov Model with state space S, initial probabilities π_i of being in state I and transition probabilities a_{ij} of transitioning from state i to state j. let us observe output $y_1,\ldots..y_T$. The most likely state sequence $x_1,\ldots,x_T$ that produces the observations is given by the recurrence relations:[2]

$$V_{1,k}=P(y_1 | k).\pi_k$$
$$V_{t,k}= P(y_t | k).\ \max_{x\varepsilon s}\ (a_{x,k}\ .V_{t-1,x})$$

Here $V_{t,k}$ is the probability of the most probable state sequence responsible for the first *t observations* that has *k* as its final state. The Viterbi path can be retrieved by saving back pointers that remember which state *x* was used in the second equation. Let $P_{tr}(k,t)$ be the function that returns the value of *x* used to compute $V_{t,k}$
If *t>1,* or *k.*
If *t*=1. Then:

$$x_T = \text{argmax}_{x\varepsilon s}(v_{T,x})$$
$$x_{t-1}= \text{Ptr}(x_t,t)$$

Here we're using the standard definition of arg max.
The complexity of this algorithm is $O(T\times|s|^2)$.

Forward backward algorithm [8]
The forward backward estimation algorithm is used to train its parameters and to find log likelihood of voice sample. It is used to estimate the unidentified parameters of HMM. It is used to compute the maximum likelihoods and posterior mode estimate for the parameters for HMM in training process. Here we want to find P(O|λ), given the observation sequence $O = O_1, O_2, O_3, \ldots\ldots, O_T$.

Forward Algorithm
The forward variable $\alpha_t(i)$ is defined as

$\alpha_t(i) = P(o_1,o_2,\ldots ,o_t,q_T = i|\lambda)$ i.e. the probability of the partial observation sequence (until time t) and state i at time t, given the model λ. αt(i) is inductively computed by following steps:
Initialization:

$\alpha_1(\iota) = \pi i\ \beta i\ (o1),1 \le \iota \le N$

Induction:

$\alpha_{t+1}(j) = [\sum\alpha_t\ (i)a_{ij}]\ b_j(\alpha_{t+1}),1 \le t \le T\ -1$

Termination:

$P(O|\lambda) = \Sigma\alpha t\ (i)$

Finally the required P(O| λ) is sum of the terminal forward variables α_T (i), this is true because

$\alpha_T\ (i) = P(O_1,O_2, \cdot\cdot\cdot ,O_T\ , q_T = S_i|\ \lambda)$

Si is the state at time t. There are N possible states Si ($1 \le i \le N$), at time t.

Backward Algorithm
The backward variable $\beta_t(i)$ is defined as:

$\beta_t(i) = P(o_{t+1},o_{t+2},\ldots,o_T,q_T = i|\lambda)$ i.e. the probability of the partial observation sequence from t+1 to the end, given the state i at time t and the model λ. β(i) is inductively solved as follows:
Initialization:

$\beta_t\ (i) = 1,1 \le i \le N$

Induction:

$\beta_t\ (i)= \Sigma\ a_{ij}b_jb_j\ (Ot+1)\ \beta_t\ +1(\ j)$ where t = T −1, T – 2…1,1 ≤i≤ N

Combining Forward and Backward variables, we get:

$P(O|\lambda) = \Sigma\alpha t\ (i)\beta_t\ (i)\ ,1 \le t \le T$

V. SPEECH RECOGNITION USING HMM [2]

Speech signal primarily consists of words or message being spoken. Area of speech recognition is concerned with determining the underlying meaning in the utterance. Success in speech recognition depends on extracting and modeling the speech dependent characteristics which can effectively

distinguish one word from another. The speech recognition system has four stages as shown in Fig. 1 [8]

i. Feature extraction
ii. Pattern training
iii. Pattern Matching
iv. Decision logic

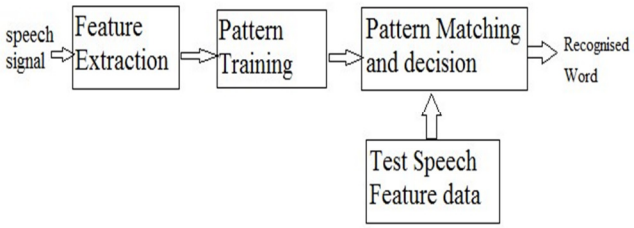

Fig. 1: Speech Recognition System[8]

Speech recognition consists of two main modules, feature extraction and pattern matching.

VI. Feature extraction

Feature extraction converts speech waveform to some type of representation for further analysis and processing, this extracted output is known as feature vector. The process of converting voice signal to feature vector is done by signal-processing front end module. As shown in above block diagram input to front-end is noise free voice sample and output of it is feature vector. Following are the few method for implementing front-end [2].

1 MFCC (Mel-Frequency Cepstral Coefficient)

This module is used to convert the speech waveform to some type of parametric representation. MFCC is used to extract the unique features of speech samples. The MFCC technique makes use of two types of filters, namely, linearly spaced filters and logarithmically spaced filters.

2 LPC (Linear Predictive Coding):

Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model.
Two types of HMMs can be used for recognition

VII. HMMs with Discrete probability distribution

This HMM modeling method is used for a process that has a discrete observation sequence. These observations could be the outcome indices of Vector Quantization technique (VQ).

VIII. HMMs with Continuous probability distribution:

It is more complex method to develop improved HMM models of the speech signals. This method needs more memory than discrete probability distribution to represent the model parameters but does not suffer from the distortion problem. On the other hand it needs more deliberate techniques to initialize the model as it might diverge easily with randomly selected initial parameters.

IX. Pattern Matching:

In pattern matching, the extracted feature vector from unknown voice sample is scored against acoustic model, the model with maximum score is selected, and its output is considered as recognized word.

X. CONCLUSION:

In this paper, we have discussed the speech recognition problem by using HMM and various frameworks such as Viterbi algorithm, forward algorithm, backward algorithm and the forward-backward estimation algorithm. We have discussed HMM in detail and found that the problems related to HMM that is evaluation problem, decoding problem and learning problem is solved by Viterbi and Backward-Forward Estimation techniques. Speech recognition models can also be implemented using various other approaches but HMMs are seen to yield maximum fruitful result. Viterbi algorithm is able to derive the maximum probability without making any assumption.

References

[1]. D.B. Paul,.Speech Recognition Using Hidden Markov Models,The Lincoln Laboratory Journal, Volume 3, Number 1,1990
[2]. Nirav S. Uchat,Hidden Markov Model and Speech Recognition,Department of Computer Science and Engineering Indian Institute of Technology, Bombay Mumbai
[3]. Bhupinder Singh, NehaKapur, Puneet Kaur,Speech Recognition with Hidden Markov Model: A Review ,International Journal of Advanced Research in Computer Science and Software Engineering, IGCE Abhipur, Mohali (Pb.), India, Volume 2, Issue 3, March 2012
[4]. Mikael Nilsson Marcus Ejnarsson,Speech Recognition using Hidden Markov Model performance evaluation in noisy environment,Department of Telecommunications and Signal Processing Blekinge Institute of Technology March, 2002
[5]. Balwant A.Sonkamble and Dr. D. D. Doye,Hidden Markov Model for Speech Recognition

[6]. Using Modified Forward-Backward Re-estimation Algorithm ,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2, July 2012 ISSN (Online): 1694-0814

[7]. Mr. Sanjay Bhardwaj, Mr. Sunil Pathania, Mr. Rajesh Akela,Speech Recognition using Hidden Markov Model and Viterbi Algorithm ,International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 5, May 2015

[8]. Waleed Abdulla, Nikola Kasabov,The Concepts of Hidden Markov Model in Speech Recognition,The Information Science Discussion Paper Series Number 99/09 May 1999 ISSN 1177-455X

[9]. Ms. Rupali S Chavan1, Dr. Ganesh. S Sable2, "An Overview of Speech Recognition Using HMM, An Overview of Speech Recognition Using HMM, International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology, IJCSMC, Vol. 2, Issue. 6, June 2013, pg.233 – 238

