Applied Mathematical Sciences, Vol. 5, 2011, no. 79, 3943 - 3948

Speech Recognition Using Hidden Markov Model

A. Srinivasan

Department of Electronics and Communication Engineering
Srinivasa Ramanujan Centre
SASTRA University, Kumbakonam-612001, India
asrinivasan78@yahoo.com

Abstract

Hidden Markov Models (HMMs) are widely used in pattern recognition applications, most notably speech recognition. Speech samples are recorded using a wave surfer tool. Wave surfer is a simple but powerful interface. The sound can be visualized and analyzed in several ways with the help of this tool. The recorded signal (test data) is compared with the original signal (trained data) using Hidden Markov Model algorithms. This speech recognition is simulated in Matlab.

Mathematics Subject Classification: 94A05, 94A12

Keywords: Speech Recognition, Wave surfer, Vector Quantization, HMM and MATLAB

1 Introduction

Stochastic processes that are usually presented as having a finite set of states, but which, in another sense, may have an infinite number of states. These processes are known variously as Hidden Markov Models (HMMs), functions of a Markov Chain, or stochastic finite automata, all of which are essentially equivalent. HMMs are used most widely and will be used here.

A process in the HMM class can be described as a finite-state Markov Chain with a memory less output process which produces symbols in a finite alphabet. This is the sense in which these processes have finitely many states. However, from the perspective of an observer who knows the parameters of some representation of the process and is able to observe the output symbols but not the internal states, things look different. For some processes there are infinitely many distinct states of such an observer's knowledge about the status of the process. This knowledge is

defined in terms of conditional distributions on future symbols. This is the sense in which there can be infinitely many states. These states are more relevant than the original finite set of states to the study of the process, since they allow for optimal prediction.

2 Analysis using Wave Surfer

The standard speech analysis such as waveform, Spectrogram, Pitch, and Power panes are analyzed. Magnitude and frequency comparison of 3 male and 3 female speakers is shown

Sl.No.	Frequency (Hz)	*F1	F2	F3	**M1	M2	M3
		dB	dB	dB	dB	dB	dB
1	15.625	-20.47	-20.98	-19.67	-21.02	-21.86	-20.71
2	140.625	-32.68	-32.73	-32.01	-32.94	-33.13	-32.75
3	390.625	-36.63	-36.74	-36.13	-36.94	-37.14	-37.01
4	640.625	-41.36	-34.99	-40.97	-41.48	-41.92	-40.99
5	1015.625	-51.20	-47.91	-51.00	-51.90	-52.65	-50.90

* Female, ** Male

Table 1: Magnitude and frequency comparison

3 Vector quantization

Vector quantization is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its center called a code word. The collection of all code words is called a code book. Vector quantization (VQ) is a lossy data compression method based on principle of block coding. It is a fixed-to-fixed length algorithm. VQ may be thought as an approximation.

The technique of VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the speaker specific features. By means of VQ, storing every single vector that we generate from the training is impossible. Fig. 1 shows a conceptual diagram to illustrate this recognition process in the figure, only two speakers and two dimensions of the acoustic space are shown. The circles refer to the acoustic vectors from the speaker 1 while the triangles are from the speaker 2. In the training phase, using the clustering algorithm described in a speaker-specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors. The result code words (centroids) are shown by black circles and black triangles for speaker 1 and 2, respectively.

The distance from a vector to the closest codeword of a codebook is called a VQ-distortion. In the recognition phase, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the total VQ distortion is computed. The speaker corresponding to the VQ codebook with smallest total distortion is identified as the speaker of the input utterance.

By using these training data features are clustered to form a codebook for each speaker. In the recognition stage, the data from the tested speaker is compared to the codebook of each speaker and measure the difference. These differences are then use to make the recognition decision.

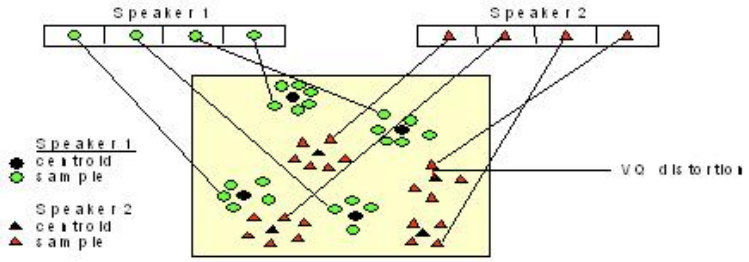

Fig 1: VQ Classification Model

Mel frequency Cepstral Coefficients are coefficients that represent audio based on perception. This coefficient has a great success in speaker recognition application. It is derived from the Fourier Transform of the audio clip. In this technique the frequency bands are positioned logarithmically, whereas in the Fourier Transform the frequency bands are not positioned logarithmically. As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system. These coefficients allow better processing of data. In the Mel Frequency Cepstral Coefficients the calculation of the Mel Cepstrum is same as the real Cepstrum except the Mel Cepstrum's frequency scale is warped to keep up a correspondence to the Mel scale.

4 Speech Recognition

The typical use of HMMs in speech recognition is not very different from the traditional pattern matching paradigm. Successful application of HMM methods usually involves the following steps

1. Define a set of L sound classes for modeling, such as phonemes or words; call the sound classes $V = \{v_1, v_2, ..., v_L\}$.

2. For each class, collect a sizable set (the training set) of labeled utterances that are known to be in the class.

3. Based on each training set, solve the estimation problem to obtain a "best" model λ_i for each class v_i (i = 1, 2, . . . , L).

4. During recognition, evaluate $P(O/\lambda)$ (i = 1, 2, . . . , L) for the unknown utterance observation and identify the speech that produced O as class v_j

5 Experimental Result

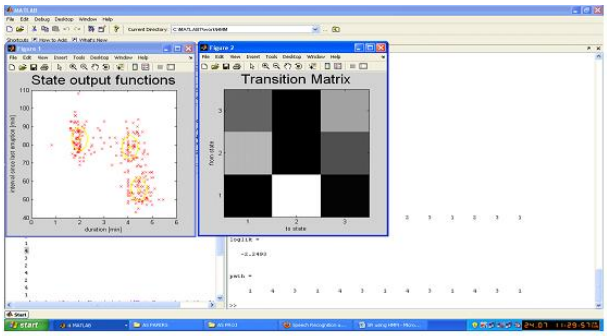

Fig. 2 : Viterbi path using Vector Quantization

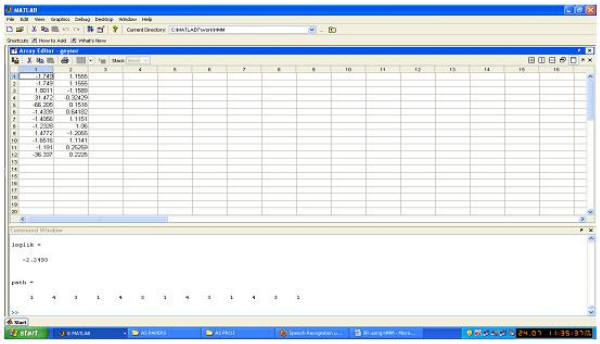

Fig. 3: Geyser discrete values of test data

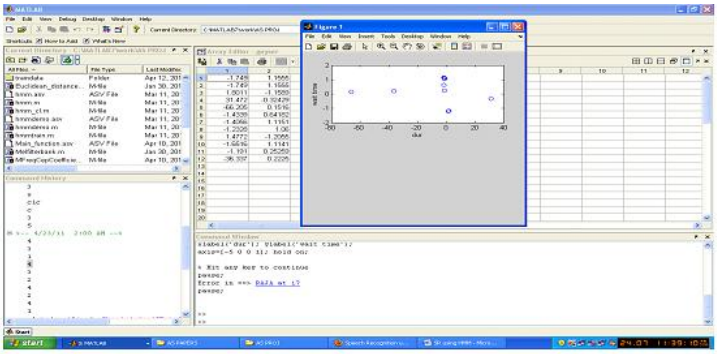

Fig. 4: Trained HMM data

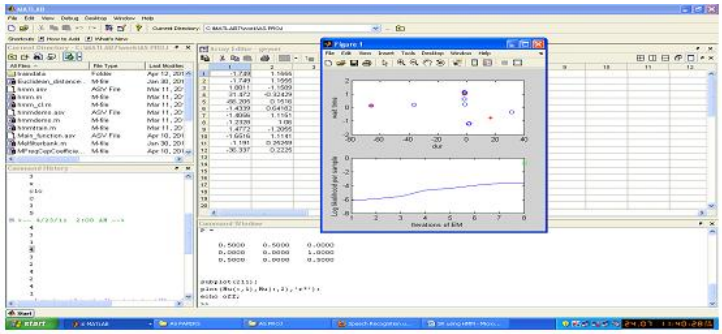

Fig. 5: Speech recognized output (Red spots show recognized output)

6 Conclusion

Speaker Recognition using Hidden Markov Model which works well for 'n' users. On the training set, hundred percentage recognition was achieved. The whole performance of the recognizer was good and it worked efficient in noisy environment also. However, the performance of the system can be improved if we have more training samples and it will be compare with the Neural networks algorithms.

References

[1] B. S. Atal, M. R. Schroeder, and V. Stover, Voice-Excited Predictive Coding Systetm for Low Bit-Rate Transmission of Speech, *Proc. ICC*, pp.30-37 to 30-40, 1975.

[2] Daniel Jurafsky, James H. Martin, Speech and Language Processing, *Pearson education, (ISBN 8178085941), 2002.*

[3] Harold F. Schiffman, A Reference Grammar of Spoken Tamil, *Cambridge University Press* (ISBN-10: 0521027527), 2006.

[4] B. H. Juang; L. R. Rabiner, Hidden Markov Models for Speech Recognition, *Technometrics*, **Vol. 33**, No. 3. (Aug., 1991), pp. 251-272.

[5] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, *Prentice- Hall, Englewood Cliffs*, NJ, 1978.

[6] J. Srinonchat, New Technique to Reduce Bit Rate of LPC- 10 Speech Coder, *IEEE Transaction on Audio and Speech Language processing*, Sweden, Sep 2006.

[7] G.F. Sudha and S. Karthik, Improved LPC Vocoder using Instantaneous Pitch Estimation Method, *International Journal of Wireless Networks and communications*, **Volume 1, Number 1** (2009), pp. 43-54.

[8] R. Thangarajan, A.M. Natarajan, M. Selvam, Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language, *WSEAS Transactions on Signal Processing*, **Issue 3, Volume 4**, March 2008.

[9] A.Srinivasan, K.Srinivasa Rao, D.Narasimhan, K.Kannan, Speech processing of the letter 'zha' in Tamil Language with LPC, *Contemporary Engineering Sciences*, **Vol. 2**, 2009, no. 10, 497 - 505.

[10] A.Srinivasan, K.Srinivasa Rao, D.Narasimhan, K.Kannan, Speech Recognition of the letter 'zha' in Tamil Language using HMM, *International Journal of Engineering Sciences and Technology*, **Vol. 1(2)**, 2009, 67 - 72.

[11] A.Srinivasan, G.Raja Krishnamurthy and G.Raghavan, Speaker identification and Verification using Vector quantization and Mel frequency Cepstral Coefficients, *(Communicated).*

[12] Wavesurfer Tool: http://mac.softpedia.com/get/Audio/WaveSurfer.shtml

Received: May, 2011