S13.3

Text-Dependent Speaker Identification Using Circular Hidden Markov Models

Yuan-Cheng Zheng Bao-Zong Yuan

Institute of Information Science
Northern Jiaotong University
Beijing, P.R.China

ABSTRACT

In this paper an approach to text-dependent speaker identification is presented in which a particular class of Markov probabilistic models, called "circular" hidden Markov models (CHMMs) by us, are first applied. CHMMs are quite different from "left-to-right" HMMs in many properties and more appropriate for speaker identification than the latter. For each person of the system, a distinct reference CHMM is produced using the Baum's forward and backward algorithm. Classification can be made by selecting the model with the highest probability as the speaker identification system output. The preliminary testing that has been done on a set of ten speakers indicates a performance of about 94 percent speaker recognition accuracy.

INTRODUCTION

Many speaker recognition systems utilize templates of averaged parameters rather than the full time sequence of parameters, since acoustic cues to speaker's identity are spread through each of his utterances. The simplest statistical approach conceptually (although not computationally) takes long term averages of speech parameters over all available data from utterances [1-2]. Stochastic modeling has also been applied to spesaker recognition with various degrees of success.

Hidden Markov models (HMMs) has recently been found to be a suitable class of stochastic models for speech recognition [3]. The probabilistic function of a Markov chain can contain information about the inherent nonstationariness of speech signal . Following is discussed the method of the application of HMMs to text-dependent speaker identification. As far as we know, it is the first attempt to utilize HMMs in the field of speaker identification (SI). A particular class of HMMs, which we call "circular" hidden Markov models (CHMMs), are found to be quite appropriate for speaker identification.

In the next section we will give a brief overview of hidden Markov models. Then the properties and the application to speaker identification of circular hidden Markov models are discussed in detail. At the last section the results of experiments are shown.

BRIEF OVERVIEW OF HIDDEN MARKOV MODELS

Let $S=\{1,2,\ldots,N\}$ denote the state space of a first order, homogeneous Markov chain, $S(t)$ the state of the Markov chain at time t, $S(t) \in S$, $t=1,2,\ldots,T$, and $C=\{C1,C2,\ldots,Cm\}$ the codebook of a vector quantizer [5-6]. Given the observation sentence $O=o_1 o_2 \ldots o_T, o_t \in C$, $t=1,2,\ldots,T$, the HMM can be specified in terms of the probability of an initial state distribution $v(i)=$ prob$\{S(0)=i\}$, $i=1,2,\ldots,N$, the probability of state transition $a_{ij}=$prob$\{S(t+1)=j/S(t)=i\}$, $i,j=1,2,\ldots,N$, and the probability of producing an obsrevation $b_{jk}=b_j(o_t)=$prob$\{o_t=Ck$, at time $t/S(t)=j\}$, $j=1,2,\ldots,N$, $k=1,2,\ldots,m$, $t=1,2,\ldots,T$. For training sequences $O^1,O^2,\ldots,O^L$, the maximum reestimates of the HMM parameters $\{a_{ij};i,j=1,2,\ldots,N\}$ and $\{b_{jk};j=1,2,\ldots,N,\ k=1,2,\ldots,m\}$ can be computed recursively from the Baum-Welch algorithm [3]:

$$a_{ij}^{n+1}=\frac{\sum_{l=1}^{L}\sum_{t=1}^{Tl-1}\alpha_t^l(i)\,a_{ij}^n\,b_j^n(o_{t+1}^l)\,\beta_{t+1}^l(j)}{\sum_{l=1}^{L}\sum_{t=1}^{Tl-1}\alpha_t^l(i)\,\beta_t^l(i)} \qquad (1)$$

$$i,j=1,2,\ldots,N$$

CH2561-9/88/0000-0580 $1.00

$$b_{ik}^{n+1}=\frac{\sum_{l=1}^{L}\sum_{t\ni o_t=C_k}\alpha_t^l(i)\beta_t^l(i)}{\sum_{l=1}^{L}\sum_{t=1}^{T_l}\alpha_t^l(i)\beta_t^l(i)} \qquad (2)$$

$$i=1,2,\ldots,N,k=1,2,\ldots,m$$

where $\alpha_t(i)$ and $\beta_t(i)$ are respectively the forward and backward probabilities of producing observation o_t and can be calculated in the recurrences:

$$\alpha_t(j)=\text{prob}\{o_1 o_2 \ldots o_t \text{ and } S(t)=j/\text{HMM}\}$$

$$=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(o_t) \qquad (3)$$

$$1\leq t\leq T-1$$

$$\beta_t(i)=\text{prob}\{o_{t+1}\ o_{t+2}\ldots o_T\ /S(t)=i \text{ and HMM}\}$$

$$=\sum_{j=1}^{N}a_{ij}b_j(o_{t+1})\beta_{t+1}(j) \qquad (4)$$

$$T-1\geq t\geq 1$$

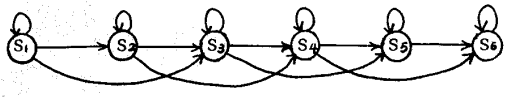
The probability of generating the observation senternce O with a HMM can be computed from the Baum algorithm[3]:

$$P=\text{prob}\{O/\text{HMM}\} \qquad (5)$$

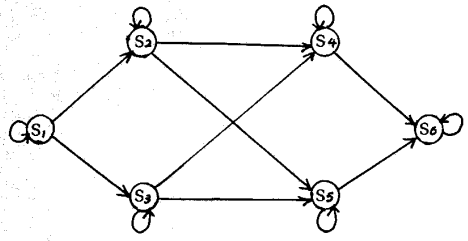
$$=\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(i)$$

for any t such that $1\leq t\leq T$.

Fig.1 shows two typical kinds of structures of left-to-right HMMs presented in [3]. Left-to-right HMM have one absorbing state at which once the Markov chain arrives the underlying Markov chain can not leaves that. Left-to-rihgt HMM has been found to be appropriate for isolated word recognition [3].



a. A serial Markov chain.



b. A parallel Markov chain.

Fig.1 The structures of two typical kinds of hidden Markov models.

CIRCULAR HIDDEN MARKOV MODELS

A. Circular Hidden Markov Models

Circular hidden Markov models are another special class of HMMs, which are different from left-to-right HMMs in structures as shown in Fig.2 and have many particular properties :

1) The first observation can be produced while the underlying Markov chain is in any state, which depends on the initial state distribution obtained by training the model.

2) The underlying Markov chain has no final or absorbing state.

3) Once the Markov chain leaves a state, that state can be revisited only at the next time.

4) Since the Markov chain has no absorbing state, the corresponding HMM can be trained by an as long training sequence as you wish.

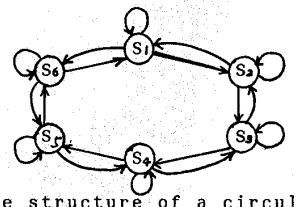


Fig.2 The structure of a circular Markov chain

B.Application of CHMMs to SI

It is assumed that the parameters in the feature set [$\{a_{ij}\}$, $\{b_{jk}\}$, $\{v(i)\}$; i,j=1, 2,...,N,k=1, 2,...m] are determined by the configuration of the vocol cord and vocol tract of a speaker and represent his identity. We can imagine that the vocol configuration is in one of a finite number of articulatory states at any time and that the probabilistic state transition law of the underlying circular Markov chain reflects the variation in target vocol tract positions for the dynamic aspect of speech, such as speaking style.

Speaker identification using CHMMs utilizes two phases: 1) a distinct CHMM needs to be trained as a reference model for each of persons known to the system, and 2) the classification of speakers consists of computing the probabilities of generating the test utterance with all reference models and selecting the model giving the highest probability as the system output.

Since a long training sequence can be used to train CHMMs, we let L=1 in the Baum-Welch reestimation fomulas 1 and 2.

For our system, we have chosen to start to train a CHMM with the following choices of the initial elements of the parameters of a CHMM:

$$v(i) = 1/N \qquad 1 \leqslant i \leqslant N. \tag{6}$$

$$\alpha_1(i) = v(i) b_i(o_1) \qquad 1 \leqslant i \leqslant N. \tag{7}$$

$$a'_{ij} = \begin{cases} 1/3, & i=1 \text{ and } j=1,2,N, \\ 1/3, & 2 \leqslant i \leqslant N-1 \text{ and } i-1 \leqslant j \leqslant i+1, \\ 1/3, & i=N \text{ and } j=1,N-1,N, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

$$b'_{jk} = 1/M \qquad 1 \leqslant j \leqslant N \text{ and } 1 \leqslant k \leqslant M. \tag{9}$$

$$\beta_T(j) = 1/N \qquad 1 \leqslant j \leqslant N. \tag{10}$$

C.Defferences Between LTR HMM and CHMM

In the structures of left-to-right HMM (LTR HMM), the absorbing state governs the fact that the rest of a single, long-observation sequence provides no further information about earlier states once the underlying Markov chain reaches the absorbing state. It is required for isolated word recognition using left-to-right HMMs that several independent utterances of the same word must be uesd for training models [3]. In the case of speaker identification, it is true that a Markov chain should be able to revisit the earlier states, because the states of a HMM reflect the vocal organic configuration of a speaker and the variations of vocal configuration may repeat in pronunciation. On the other hand,both training and testing sentences are usually more than 30s long for good performance of speaker identification.

Obviously, it is irrational and unconvenient to utilize left-to-right HMMs having one absorbing state for speaker identification. Circular HMM is presented with the object of resolving the problems mentioned above for using HMMs for speaker identification. Experiments show that CHMMs are more appropriate for speaker identification than left-to-right HMMs.

EXPERIMENTS AND RESULTS

The entire speaker identification system based on CHMMs has been evaluated on a set of 10-person population (6 male and 4 female). The speech was first low-pass filted to 3.4KHz and then sampled at the rate of 10KHz. After the end point detection, the speech signal was segmented into 15ms, 12-order LPC analysis was done on each frame,and a clustering algorithm [6] was used on these LPC coefficient vectors to obtain a codebook consisting of 64 codewords. The orders of the state transition matrix $[a_{ij}]$ of an underlying circular Marvok chain and a stochastic matrix $[b_{jk}]$ was 6x6 and 6x64 respectively.

For training, 7 independent chinese sentences were spoken by each of the 10 persons at a time and a CHMM representing the identity of a speaker was trained for each of the 10 persons. For testing, an uttrance of one of the same sentences was obtained from one of the population at a time. Computing the probability of generating the uttrance with each CHMM, the model with the highest probability was chosen as the output of the speaker identification system. For the 7-chinese sentence test set, the averaged recognition accuracy was found to be 93.7 percent for speaker identification system using CHMMs, 88.7 percent for the system using serial left-to-right HMMs, and 90 percent for the system using parallel left-to-right HMMs.

The preliminent experiments show that the application of the CHMMs presnted here to speaker identification is feasible and CHMMs is more appropriate for speaker identification than left-to-right HMMs.

References

1. H.Hollien, and W.Majewski, " Speaker identification by long-term spectra under normal and distorted speech conditions," J.acoust.Soc.Am., Vol.62, pp.975-980, 1977.
2. S.Furui, " Comparison of speaker recognition metheods using statistical features and dynamic festures," IEEE Trans. ASSP, Vol.ASSP-29, pp. 342-350, 1981.
3. S.E.Levinson, L.R.Rabiner,and M.M.Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," B.S.T.J., Vol.62, No.4. Part 1, pp.1035-1074, April,1983.
4. L.R.Rabiner, S.E.Levinson,and M.M.Sondhi, "On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition," B.S.T.J., Vol.62, No.4, Part 1, pp.1075-1105, April, 1983.
5. H.Abut et al., "Vector quantization of speech," IEEE Trans. ASSP,Vol.ASSP-30, No.3, pp.423-435, 1982.
6. Y.Linde, A.Buzo, R.M.Gray, " An algorithm for vector quantizer design," IEEE Trans. COMM., Vol.COM-28, pp.84-95, Jan.1980.