PARTLY HIDDEN MARKOV MODEL AND ITS APPLICATION TO SPEECH RECOGNITION

Tetsunori Kobayashi, Junko Furuyama, Ken Masumitsu
Dept. EECE, Waseda University
Shijuku-ku, Tokyo 169-8555, Japan

Abstract

A new pattern matching method, Partly Hidden Markov Model, is proposed and applied to speech recognition.

Hidden Markov Model, which is widely used for speech recognition, can deal with only piecewise stationary stochastic process. We solved this problem by introducing the modified second order Markov Model, in which the first state is hidden and the second one is observable. In this model, not only the feature parameter observations but also the state transitions are dependent on the previous feature observation. Therefore, even the compricated transient can be modeled precisely.

Some simulational experiments showed the high potential of the proposed model. As the results of word recognition test, the error rate was reduced by 39% compared with normal HMM.

1 Introduction

In the HMM, the output probability of feature vectors in each state is unique. That means HMM basically neglect the dynamic features and deal with only piecewise stationary process. Since the dynamic features of speech must play important role in the speech hearing, it is desired for the speech recognizers to deal with them appropriately. If the dynamical features of the speech patterns are modeled more precisely, it must contribute to improve the reliabilities of likelihood and improve the performance. It may be useful for the word spotting or pruning in continuous speech recognition also.

Aiming at this point, several approach to improve HMM have been tried. Deng[1] introduced trajectory models in the HMM. Mari[2] and Ariki[3] proposed 2nd order HMM. Wellekens[4], Takahashi[5] and Digalakis[6] introduced inter-frame dependence in HMM.

In this paper, we propose another model for time series pattern matching, that is modified 2nd order Markov model. The first state of this model is hidden and the second one is observable. For this structure, not only the feature parameter observations but also the state transitions are dependent on the previous feature observation. Thus, it is possible to deal with transient process rather than piecewise stationary. We call this model Partly Hidden Markov Model (PHMM).

In the next section, PHMM is introduced in contrast to HMM. In this section, we also describe some important smoothing technique to improve the performance. In section 3 some results of simulation experiments are described to explain the merit of this model. Finally in section 4, the experimental results of speech recognition are shown.

2 Partly Hidden Markov Model

2.1 Markov Model and Hidden Markov Model

In general, the output probability of feature vector x_t, $P_t(x_t)$, is given by the conditional probability of the all past observation $x_0 x_1 \cdots x_{t-1}$

$$P_t(x_t) = Pr(x_t | x_0 x_1 \cdots x_{t-1}). \quad (1)$$

In the Markov model, the number of the sequence which is used for the condition is truncated by fixed number K.

$$P_t(x_t) = Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1}). \quad (2)$$

And, certain state S_i is uniquely given to the sequence of $x_{t-K} x_{t-K+1} \cdots x_{t-1}$. Then, the former equation become

$$P_t(x_t) = Pr(x_t | S_i). \quad (3)$$

In the Hidden Markov Model, the same representation is adopted for the output probability. However, in this case, the relation between the output sequence and state is not unique but probabilistic.

2.2 PHMM

In the proposed model, the output probability $Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1})$ is represented by second order model,

$$Pr(x_t | x_{t-K} x_{t-K+1} \cdots x_{t-1}) = Pr(x_t | S_i^f, S_j^s). \quad (4)$$

Here, state S_i^f is given to the sequence of $x_{t-K} x_{t-K+1} \cdots x_{t-2}$ and state S_j^s is given to the outout of x_{t-1}. We call S_i^f f-state (first state). And we call S_j^s s-state (second state). If both of these mappings are unique, this model is equivalent to the

Markov Model. If both of them are probabilistic, it is equivalent to the Hidden Markov Model.

In the proposed model, mapping from the sequence of $x_{t-K}x_{t-K+1}\cdots x_{t-2}$ to state S_i^f is probabilistic and the mapping from the output x_{t-1} to the state S_j^s is unique. We call this model "Partly Hidden Markov Model (PHMM)." Since the half of the conditional part of the output probability is shared in many varieties of output sequences, the number of the states S_i^f can be reduced and the complexity of the model can also be reduced. Since the output probability of x_t is conditioned by state S_j^s (that means it is conditioned by x_{t-1}), the model can deal with more complicated process than piecewise stationary.

In the proposed model, the probability that the output sequence $x_1x_2\cdots x_T$ (s-state transition is $x_0x_1x_2\cdots x_{T-1}$) comes from the model with the f-state transition $s_1s_2\cdots s_T$ is defined by following equation.

$$Ps = Pr(x_1x_2\cdots x_T s_1^f s_2^f\cdots s_T^f s_1^s s_2^s\cdots s_T^s) \quad (5)$$

Since $s_1^s s_2^s\cdots s_T^s = x_0x_1\cdots x_{T-1}$,

$$\begin{aligned} Ps &= Pr(x_1x_2\cdots x_T s_1^f s_2^f\cdots s_T^f x_0) \\ &= Pr(s_1^f s_1^s)Pr(x_1|s_1^f,s_1^s) \\ &\quad \cdot\prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f s_t^s)Pr(x_{t+1}|s_{t+1}^f s_{t+1}^s) \\ &= Pr(s_1^f,x_0)Pr(x_1|s_1^f x_0) \\ &\quad \cdot\prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f x_{t-1})Pr(x_{t+1}|s_{t+1}^f x_t) \quad (6) \end{aligned}$$

Also, since,

$$Pr(s_{t+1}^f \mid s_t^f x_{t-1}) = \frac{Pr(s_{t+1}^f|s_t^f)Pr(x_{t-1}|s_{t+1}^f s_t^f)}{Pr(x_{t-1}|s_t^f)} \quad (7)$$

$$Pr(x_{t+1}|s_{t+1}^f x_t) = \frac{Pr(x_{t+1}x_t|s_{t+1}^f)}{Pr(x_t|s_{t+1}^f)} \quad (8)$$

Eq. (6) becomes,

$$\begin{aligned} Ps &= Pr(s_1^f,x_0)Pr(x_1|s_1^f,x_0) \\ &\quad \cdot\prod_{t=1}^{T-1}\frac{Pr(s_{t+1}^f|s_t^f)Pr(x_{t-1}|s_{t+1}^f s_t^f)}{Pr(x_{t-1}|s_t^f)} \\ &\quad \cdot\frac{Pr(x_{t+1}x_t|s_{t+1}^f)}{Pr(x_t|s_{t+1}^f)}. \quad (9) \end{aligned}$$

$Pr(x_1x_2\cdots x_T)$ can be obtained by summing up Eq.(9) for all possible combination of F-state transition $s_1^f s_2^f\cdots s_T^f$.

From above discussion, it is found that PHMM can be expressed by following 5 parameters.

a_{ij} : the probability that the next f-state is S_j^f in case that the current f-state is S_i^f.

$b_i(x)$: the probability that the current S-state is x in case that the current f-state is S_i^f.

$c_{ij}(y)$: the probability that the current s-state (last output) is y in case that the current f-state is S_i^f and the next f-state is S_j^f.

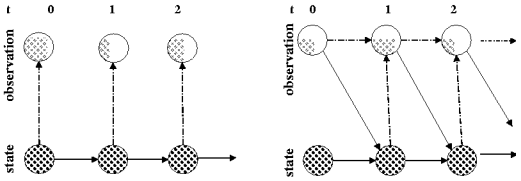
$d_i(x,y)$: the probability that the current output is x and the current s-state (last output) is y in case that the current f-state is S_i^f.

$e_i(y)$: the probability that the initial s-state is y and the initial f-state is S_i^f.

The training algorithm to get above PHMM parameters can be derived through EM algorithm or segmental K-means algorithm like HMM training algorithm. The forward algorithm and viterbi algorithm to get likelihood for PHMM can be derived by similar way to the HMM.

Figure 1 shows the characteristics of PHMM compared to the HMM. Arrows in the figure show the dependency in the observation sequence and state sequence. In HMM, state transition is dependent only on the previous state and observation is dependent only on current state. While, in PHMM, state is defined by the previous f-state and previous observation, and observation is dependent on current f-state and previous observation (current s-state).

This effect is easily found in the state transition. In HMM, the probability moving to next state is constant in a state. While, in PHMM, the state transition probability is changing according to the previous observation even in a state (See Fig.2).



(a) HMM (b) PHMM

Fig.1 Dependency of the observation sequence and the state transition sequence. Arrows in the figures show the dependency.

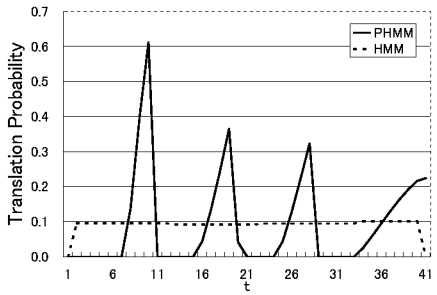


Fig.2 An example of temporal change of the transition probabilities.

2.3 Smoothing of observation probabilities

PHMM used higher order statistics than HMM. Higher order statistics is sometimes less reliable. In that case, smoothing with lower order statistics is effective. For example, it is well known that the trigram language model smoothed with bigram perform better thatn simple trigram.

Therefore, we adopted smoothing in PHMM. Namely, we use

$$\frac{Pr(x_{t+1}x_t|s^f_{t+1})^\alpha}{Pr(x_t|s^f_{t+1})^\alpha} \cdot Pr(x_{t+1}|s^f_{t+1})^{1-\alpha}. \quad (10)$$

instead of

$$\frac{Pr(x_{t+1}x_t|s^f_{t+1})}{Pr(x_t|s^f_{t+1})} \quad (11)$$

in Eq. (9)

3 Simulation

3.1 Discrimination of transient difference

In order to show the effectiveness of PHMM, two simulation experiments are performed.

In the first simulation experiment, we examine the ability of PHMM for the discrimination of two signals whose target values are the same but transients are different, one transient is piecewise stationary and the other is with gentle slope.

We select rectangler signal as the the piecewise stationary signal and cosine signal as gentle slope signal. Signals are slightly fluctuated by random sequence.

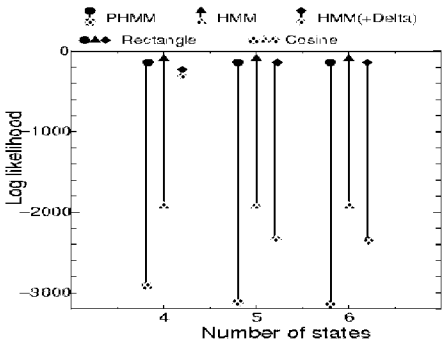
We examined the average likelihood value of each signal for each model.

Figure 3 shows the results. Horizontal axis denotes the experimental condition of the state number. Every three lines correspond to the same state number condition. Left most line of every three lines denotes the results of PHMM, middle one is HMM without delta parameters and right one is HMM with delta parameters. The top points of each line denotes the average likelihood of signal observation when we used the right signal for the model (test category is the same as training category). The bottom points of the lines denote the likelihood when we used the wrong signal (the test category is different from training category). The top points is higher the better and bottom point is lower the better. So, the long line shows the good ability of discrimination.

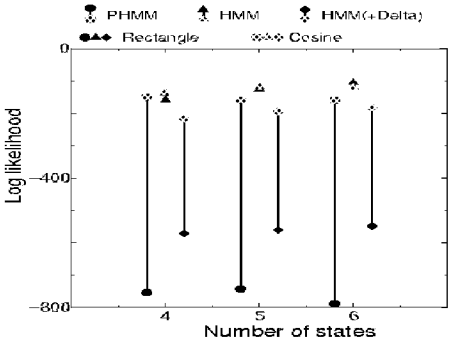
From these figures, it was shown that the piecewise stationary signals are well treated by all models, HMM and PHMM, however, cosine signals which have gentle slope can not be treated by HMM without delta parameter. Using HMM with delta, the discrimination ability is improved. However, the top position of HMM with delta is approximately 20% lower than that of PHMM. This results shows the high reliability of PHMM based likelihood values. The ability of PHMM is the best.

3.2 Discrimination of target difference from transient

In the second simulation experiment, we examine the ability of PHMM for the discrimination of two groups of signals with different targets using transient data.

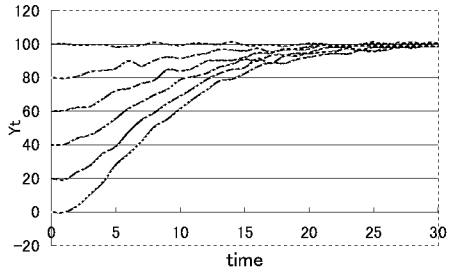


(a) Rectangle trained models.

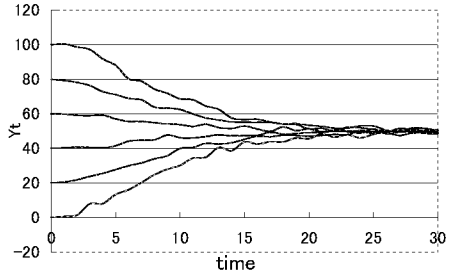


(b) Cosine trained models.

Fig.3 Average likelihood values in 1st simulation experiment.



(a) Category 1

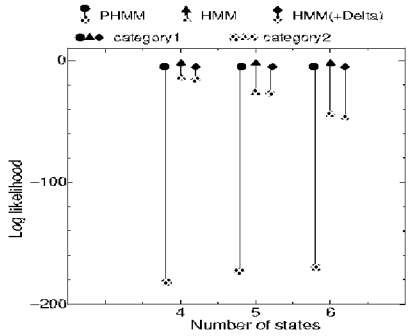


(b) Category 2

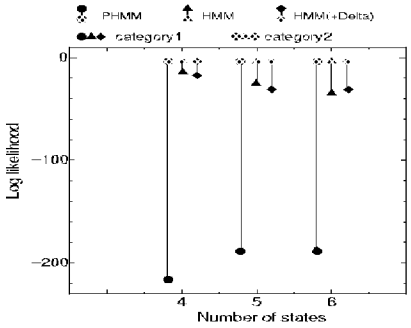
Fig.4 Examples of signals used for 2nd simulation experiment.

All signals in the two groups have the same dynamics of 2nd order critical damping. Signals in a group have wide variation of initial values but the same target. Targets of the groups are different each other (Target of the category 1 is 100 and that of category 2 is 50). These signals also fluctuated by random sequence. Figure 4 show the examples of the signals we used in this experiment.

Figure 5 shows the results. From this figure, it was found that the PHMM can discriminate the target difference from transient data in spite of wide variation of initial values. This results imply that the context independent models might perform well in PHMM-base recognizer, while context dependent models are indispensable in HMM-based one.



(a) Category 1 (target 100) trained model



(b) Category 2 (target 50) trained model

Fig.5 Average likelihood values in 2nd simulation experiment.

4 Speech Recognition Experiment

To evaluate the effectiveness of PHMM for speech recognition, word recognition experiment was done.

The task is to discriminate ATR 216 phonetically balanced word set uttered by 7 male speakers.

The word models are constructed by concatenating mono-phone models trained with 10000 sentences from JNAS speech database[7]. Distribution function of each state in models is represented by a normal distribution with full covariance. Since the JNAS has no phonetic labels and we have not developed the training algorithm of PHMM parameters from label-less database yet, we automatically labeled JNAS data using HMM and used them for the training. This is rather unfavorable process for PHMM because optimal segment boundary for HMM is not always optimal for PHMM. The microphone used for test data is different from that of training data. However, we did not used particular process to compensate them.

Figure 6 show the results of experiments. Horizontal axis denotes the smoothing parameter α in Eq. (10). The result in case of $\alpha = 0$ corresponds to HMM. The result in case of $\alpha = 1$ corresponds to pure PHMM. From this figure, it is found that the smoothing is very important to improve performance. PHMM with smoothing factor 0.8 give the best score 93.4%. This is 39% error rate reduction.

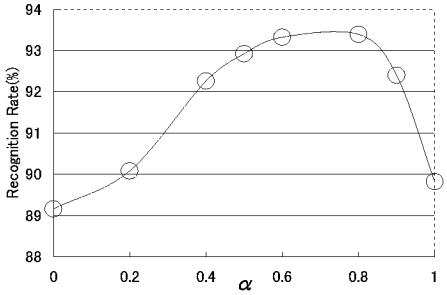


Fig.6 Results of word recognition as a function of smoothing parameter α.

5 Conclusion

A new stochastic model named PHMM is proposed and it is applied to speech recognition. As compared with HMM, PHMM improved error rate by 39%.

Since PHMM realize high reliability of likelihood values, it is expected that it suit for word spotting and pruning. Simulation experiment imply the possibility of PHMM in context independent modeling. In the next stage, we'd like to evaluate the effectiveness of PHMM in these points of view.

References

[1] L. Deng, M. Aksmanovic, "Speaker-Independent phonetic classification using Hidden Markov Models with mixture of trend functions" , IEEE trans on Speech and Audio Processing, vol. 5, pp.319-324, JULY. 1997.

[2] J-F. Mari, J-P. Haton, "Automatic word recognition based on second-oder hidden Moarkov models," IEEE Trans. on Speech and Audio Process, vol.5, n.1, Jan. 1997.

[3] Y. Ariki, "Mixture density HMMs with two-level transition", Journal of Acoustic Society Japan(E), vol.14, no.4, pp.279-280, Sep. 1993.

[4] C.J. Wellekens, "Explicit correlation in Hidden Markov Model with optimal inter-frame dependence", Proc. ICASSP87, pp.383-386, 1987.

[5] S.Takahashi, T.Matsuoka, Y.Minami and K.Shikano, "Phoneme HMMs constrained by frame correlations," Proc. ICASSP93, pp.219-222, 1993.

[6] V. Digalakis, M. Ostendorf and J.R. Roglicek, "Improvement of the stochastic segment model for phoneme recognition," Proc. DARPA Workshop on Speech and Natural Language, pp.332-338, 1989.

[7] K.Itou etal., "Design and Development of Japanese Speech Corpus for Large Vocabulary Continuous Speech Recognition Assesment," Proc. EALREW, pp.98-103, 1998.