International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Hidden Markov Models and Its Application

A. Juliet Dayana[1], V. Krishnan[2]

[1]Assistant Professor, Department of Mathematics, Marudupandiyar College, Thanjavur-613 403.

[2]Assistant Professor & Research Advisor, PG and Research Department of Mathematics, Jamal Mohamed College (Autonomous) Tiruchirappalli – 620 020.

Abstract

In this paper We attempt to carefully and methodically review the theoretical aspects of this type of statistical modeling and show how they have been applied to selected problem in machine recognition of speech.

Keywords: Markov Model, Coin Toss Model, Ball Model, Basis Model

1. INTRODUCTION

Although initially introduced and studied in the late 1960 and early 1970 statistical methods of markov source or hidden markov modeling have become increasingly popular in the last several years. There are two strong reasons why this has occurred. First the models are very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of application[1]. Second the models, when applied properly work very well in practice for several important applications.

Real-world processes generally produce observable outputs which can be characterized as signals. The signals can be discrete in nature e.g characters from a finite alphabet, quantized vectors from a codebook etc. or continuous in nature e.g speech samples, temperature measurements, music etc. The signal source can be stationary that is its statistical properties do not vary with time, or nonstationary the signal properties vary over time .The signals can be pure coming strictly from a single source or can be corrupted from other signal sources or by transmission distortions, reverberation etc.

A problem of fundamental interest is characterizing such real-world signals in terms of signal models. There are several reasons why one is interested in applying signal models. First of all a signal model can provide the basis for a theoretical description of a signal processing system which can be used to process the signal so as to provide a desired output. For example if we are interested in enhancing a speech signal corrupted by noise and transmission distortion, we can use the signal model to design a system which will optimally remove the noise and undo the transmission distortion[2]. A second reason why signal models are important is that they are potentially capable of letting us learn a great deal about the signal source i.e., the real-world process which produced the signal without having to have the source available. This property is especially important when the cost of getting signals from the actual source is high. In this case, with a good signal model, we can simulate the source and learn as much as possible via simulations. Finally, the most important reason why signal models are important is that they often work extremely well in practice, and enable us to realize important practical systems- e. g prediction systems, recognition systems, identification systems, etc., in a very efficient manner.

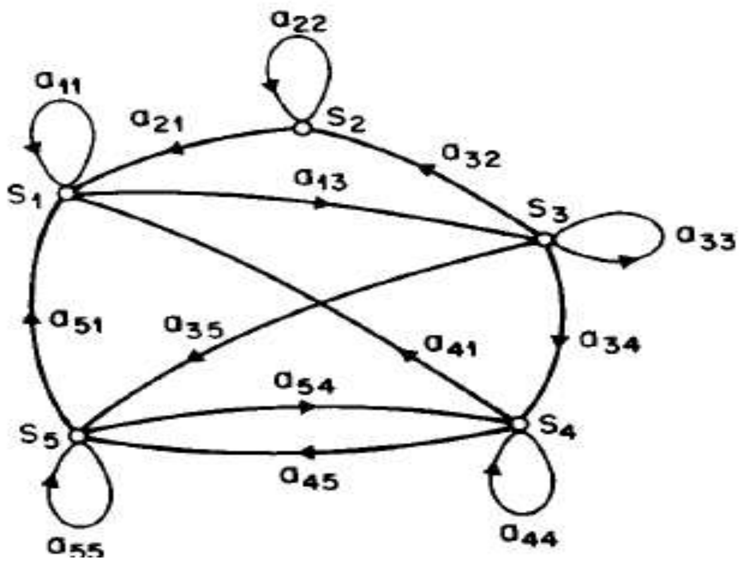
For the application of interest, namely speech processing , both deterministic and stochastic signal models have had good success. we will concern ourselves strictly with one type of stochastic signal model, namely the hidden markov model (HMM).These models are referred to as Markov source or probabilistic functions of markov chains in the communications literature[3]. We will first review the theory of markov chains and then extend the ideas to the class of hidden Markov models using several simple example. We will then focus our attention on the three fundamental problems' for HMM design, namely the evaluation of the probability of a sequence of observations given a specific HMM the determination of a best sequence of model states., and the adjustment of model parameters so as to best account for the observed signal we will show that once these the five fundamental problems are solved. We can apply HMMs to selected problems in speech recognition.

2. DISCRETE MARKOV PROCESSES

2.1 Observable Markov Model :

Consider a system which may be described at any time as being in one of a set of N distinct states, S_1 , S_2,…., S_N.,as illustrated in fig 1(a) and (b) where N =5 and N=6 for simplicity. At regularly spaced discrete times , the system undergoes a change of state possibly back to the same state according to a set of probabilities associated with the state. We denote the time instants associated with state changes as t = 1,2,…., and we denote the actual state at time t as q_t .A full probabilistic description of the above system would[4]. In general, require specification of the current state at time D, as well as all the predecessor states.

Fig. 1a , A Markov Chain with 5 states (labeled S_1 to S_5) with selected state transitions.



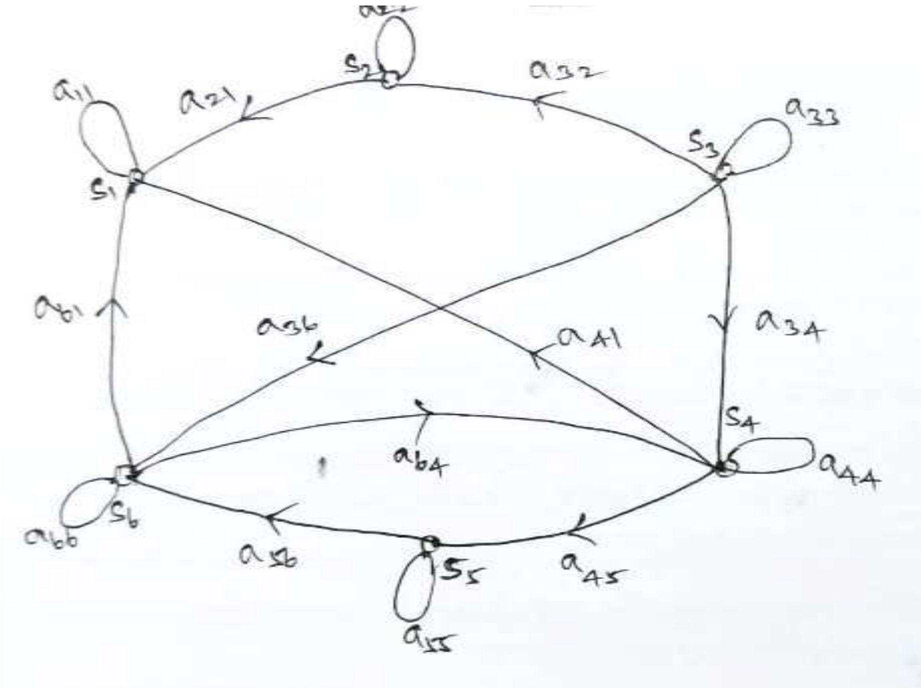


Fig.1b, A Markov Chain with 6 states (labeled S_1 to S_6) with selected state transitions.

2.2 A Simple 3- State Markov model :

Consider a simple 3-state Markov model of the weather. We assume that once a day the weather is observed as being one of the following:

State 1 : Rain or (Snow)

State 2 : Cloudy

State 3 : Sunny

We postulate that the weather oneday r is characterized by a single one of the three states above, and that the matrix A of state transition probabilities is

$$A = \{ a_{ij} \} = \begin{bmatrix} 0.2 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.1 \\ 0.5 & 0.5 & 0.6 \end{bmatrix}$$

Given that the weather on day 1 (t =1) is sunny (state 3), we can ask the question: What is the probability (according to the model) that the weather for the next 7 days will be "sun-sun-rain-sun-cloudy-sun …." t Stated more formally, we define the observation sequence O as O = { S_3 , S_3 , S_3, S_1 , S_1, S_3, S_2, S_3 } corresponding to t = 1,2,….,8, and we wish to determine to probability of O, given the model .This probability can be expressed (and evaluated) as

P(O | Model = P [S_3 , S_3 , S_3, S_1 , S_1, S_3, S_2, S_3 | Model]

= P [S_3] . P [$S_3|S_3$] . P [$S_3|S_3$] . P [$S_1|S_3$]

. P [$S_1|S_1$] . P [$S_3|S_1$] . P [$S_2|S_3$] . P [$S_3|S_2$]

= π_3 . a_{33} . a_{33}. a_{31}. a_{11}. a_{13}. a_{32} . a_{23}

= 1. (0.6)(0.6)(0.5)(0.2)(0.3)(0.5)(0.1)

= 5.4 x 10^{-4}

Where we use the notation

π_1 = P [$q_1 = S_1$], 1 ≤ i ≤ N

To denote the initial state probabilities.

Another interesting question we can ask (and answer using the model) is given that the model is in a known state[5], what is the probability it stays in that state for exactly d dayst this probability can be evaluated as the probability of the observation sequence

O ={ $\frac{S_i}{1}, \frac{S_i}{2}, \frac{S_i}{3}, \ldots, \frac{S_i}{d}, \frac{S_i}{d+1} \neq S_i$} ,

Given the model , which is

P(O | model , $q_1 = S_1$) = ($a_{ij})^{d-1}$ (1 - a_{ij}) = p_i(d) .

The quantity p_i(d) is the discrete probability density function of duration d in state i. This exponential duration density is characteristic of the state duration in a markov chain. Based on p_i(d) we can readily calculate the expected number of observations duration in a state, conditioned on starting in that state as

$\overline{d_1} = \sum_{d=1}^{\infty} dp_i$(d)

= $\sum_{d=1}^{\infty} d(a_{ii})^{d-1}$ (1 - a_{ii}) = $\frac{1}{1-a_{ii}}$

Thus the expected number of consecutive days of sunny weather, according to the model is 1/(0.4) = 2.5 ; for cloudy it is 1.67; for rain it is 1.25.

2.3 Coin Toss Models :

Assume the following scenario. you are in a room with a barrier e.g a curtain through which you cannot see what is happening. On the other side of the barrier is another person who is performing a coin or multiple coin tossing experiment. The other person will not tell you anything about what he is doing exactly; he will only tell you the result of each coin flip. Thus a sequence of hidden coin tossing experiments is performed, with the observation sequence consisting of a series of heads and talls e.g a typical observation sequence would be

O = $O_1, O_2, O_3 \ldots. O_T$

= **K K K J J J K J J K … K**

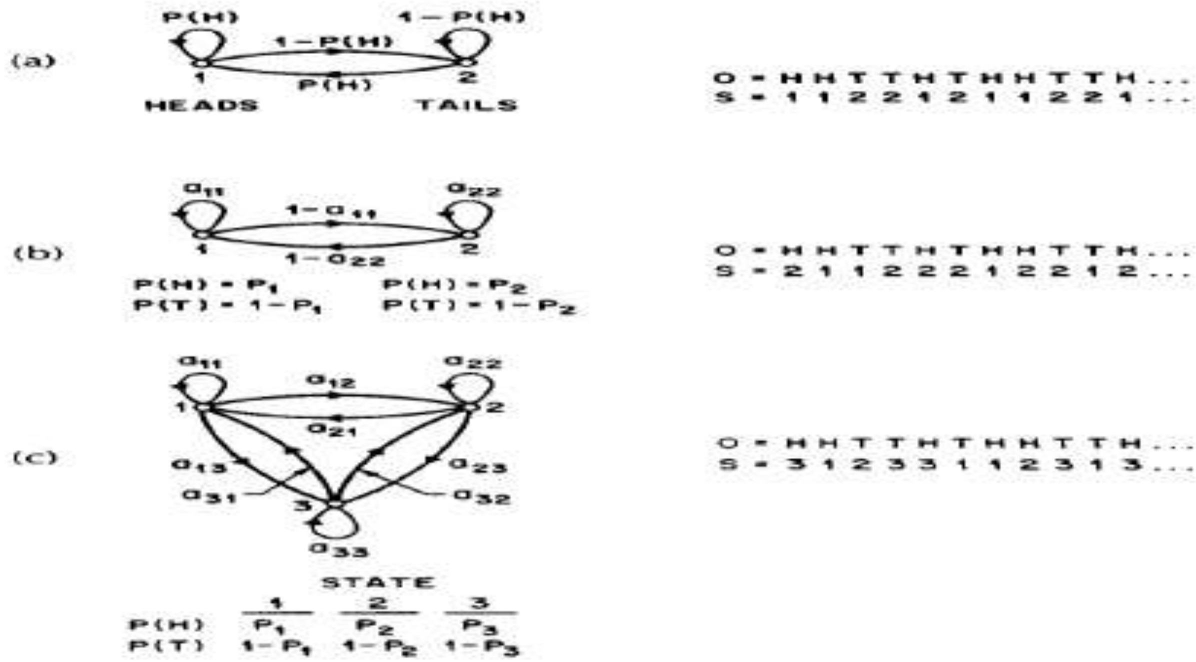
Where **K** stands for heads and 2 stands for tails.

Given the above scenario, the problem of interest is how do we build an HMM to explain the observed sequence of heads and tails[6]. The first problem one faces is deciding what the states in the model correspond to, and then deciding how many states should be in the model. One possible choice would be to assume that only a single biased coin was being tossed. In this case we could model the situation with a 2-state model where each state corresponds to a side of the coin (i.e heads or tails) This model is depicted in Fig.2(a) . In this case the markov model is observable, and the only issue for complete specification of the model would be to decide on the best value for the bias(i.e., the probability of , say, heads). Interestingly an equivalent HMM to that of Fig 2(a) would be a degenerate 1-state model, where the state corresponds to the single biased coin, and the unknown parameter is the bias of the coin.

A second form of HMM for explaining the observed sequence of coin toss outcome is given in Fig .2(b) . In this case there are 2 states in the model and each state corresponds to a different , biased , coin being tossed. Each state is characterized by a probability distribution of heads and tails, and transitions between states are characterized by a state transition matrix. The physical mechanism which account for how state transitions are selected could itself be a set of independent coin tosses, or some other probabilistic event[7].

A third form of HMM for explaining the observed sequence of coin toss outcomes is given in Fig2(c) . This model corresponds to using 3 biased coins, and choosing from among the three , based on some probabilitistic event.

Fig.2. Three possible Markov models which can account for the results of hidden coin tossing experiments.(a)1-coin model.(b)2-coins model .(C) 3-coins model.



Given the choice among the three models shown in Fig.2 for explaining the observed of heads and tails, a natural question would be which model best matches the actual observations. It should be clear that the simple 1-coin model of Fig-2(a) has only 1 unknown parameters, the 2-coin model of Fig.2(b) has 4 unknown parameters; and the 3-coin model of Fig.2(c) has 9 unknown parameters. Thus , with the greater degree of freedom , the larger HMMs would seem to inherently be more capable of modeling a series of coin tossing experiments than would equivalently smaller models[8]. Although this is theoretically true, we will see later in this paper that practical considerations impose some strong limitations on the size of models that we can consider. Furthermore, it might just be the case that only a single coin is being tossed. Then using the 3-coin model of Fig.2(c) would be inappropriate, since the actual physical event would not correspond to the model being used-i.e., we would be using an underspecified system.

2.4 The Urn and Ball Model :

To extend the ideas of the HMM to a somewhat more complicated situation, consider the urn and ball system of Fig.3. We assume that there are N(large) glass urns in a room. Within each urn there are a large number of colored balls. We assume there are M distinct colors of the balls[9]. The physical process for obtaining observations is as follows. A genie is in the room, and according to some random process, he choose an initial urn. From this urn, a ball is chosen at random, and its color is recorded as the observation . The ball is then replaced in the urn from which it was selected. A new urn is then selected.

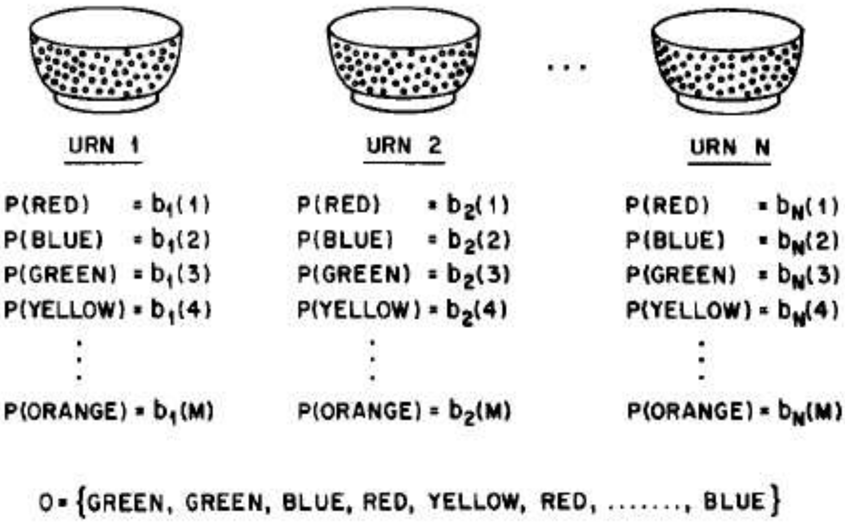


Fig. 3. An *N*-state urn and ball model which illustrates the general case of a discrete symbol HMM.

According to the random selection process associated with the current urn, and the ball selection process is repeated. This entire process generates a finite observation sequence of colors, which we would like to model as the observable output of an HMM.

It should be obvious that the simplest HMM that corresponds to the urn and ball process is one in which each state corresponds to a specific urn, and for which a color probability is defined for each state[10]. The choice of urns is dictated by the state transition matrix of the HMM.

2.5 A Basis Model :

Until now, we have only considered the special case of ergodic or fully connected HMMs in which every state of the model could be reached (in a single step) from every other state of the model. (Strictly speaking, an ergodic model has the property that every state can be reached from every other state in a finite number of steps[11]. As shown in fig 4(a) , for an N=4 state model, this type of model has the property that every a_{ij} coefficient is positive. Hence for the example of Fig.4a we have

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

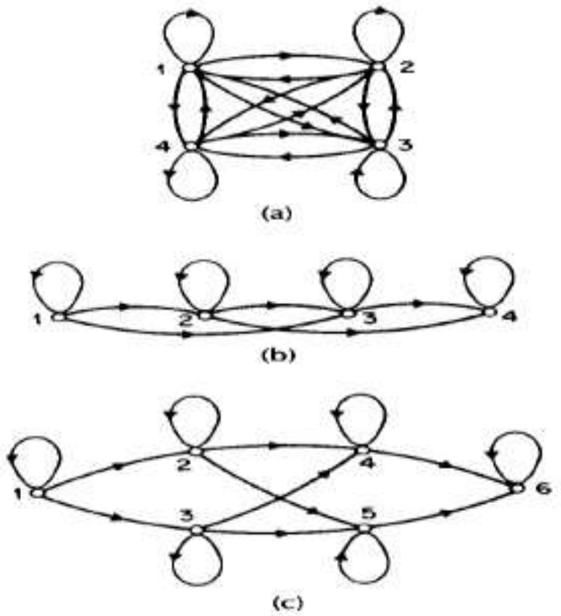


Fig. 4 Illustration of 3 distinct types of HMMs. (a) A 4-state ergodic model. (b) A 4-state left-right model. (c) A 6-state parallel path left-right model.

For some applications, in particular those to be discussed later in this paper ,other types of HMMs have been found to account for observed properties of the signal being modeled better then the standard ergodic model. One such model is shown in Fig.4(b).This model is called a left-right model of a bakis model because the underlying state sequence associated with the model has the property that as time increases the state index increases for stays the same. i.e the states proceed from left to right[12]. Clearly the left-right type of HMM has the desirable property that it can readily model signals whose properties change over time e.g. speech. The fundamental property of all left-right HMMs is that the state transition coefficients have the property.

$a_{ij} = 0 , j < i$

i.e. no transitions are allowed to states whose indices are lower than the current state. Furthermore , the initial state probabilities have the property

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

Since the state sequence must begin in state 1 (and end in state N) often with left-right models, additional constraints are placed on the state transition coefficients to make sure that large changes in state indices do not occur hence a constraint of the form

$a_{ij} = 0 , j > i + \Delta$

is often used. In particular , for the example of Fig.4(b),the value of Δ is 2, i.e no jumps of more than 2 states are allowed. The form of the state transition matrix for the example of Fig.4(b) is thus

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \end{pmatrix}$$

$0 \quad 0 \quad 0 \quad a_{44}$

A=

It should be clear that , for the last state in a left-right model, that the state transition coefficients are specified as

a_{NN} = 1

a_{Ni} = 1 , I < N

Although we have dichotomized HMMs into ergodic and left-right models, there are many possible variations and combinations possible[13]. By way of example, Fig 4(c) shows a cross-coupled connection of two parallel left-right HMMs, Strictly speaking, this model is left-right model (it obeys all the a_{ij} constrints); however, it can be seen that it has certain flwxibility not present in a strict left-right model (i.e., one without parallel paths.).

It should be clear that the imposition of the constraints of the left-right model, or those of the constrained jump model, essentially have no effect on the reestimation procedure. This is the case because any HMM parameter set to zero initially, will remain at zero throughout the reestimation procedure.

CONCLUSION

We have attempted to present the theory of hidden Markov models from the simplest concepts of (discrete Markov chains) to the most sophisticated models in (variable duration, continuous density models). It has been our purpose to focus on physical explanations of the basic mathematics. Hence we have avoided long, drawn out proofs and /or derivations of the key results, and concentrated primarily on trying to interpret the meaning of the math, and how it could be implemented in practice in real world systems. We have also attempted to illustrate some applications of the theory of HMMs to simple problems in speech recognition, and pointed out how the techniques could be applied to more advanced speech recognition problems.

REFERENCES

1. Edgar R C, Sj¨olander K. COACH: profile-profile alignment of protein families using hidden Markov models. *Bioinformatics.* 2004;**20**:1309–1318.
2. L¨oytynoja A, Milinkovitch M C. A hidden Markov model for progressive multiple alignment. *Bioinformatics.* 2003;**19**:1505–1513.
3. Liang K C, Wang X, Anastassiou D. Bayesian basecalling for DNA sequence analysis using hidden Markov models. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2007;**4**:430–440.
4. Lottaz C, Iseli C, Jongeneel C V, Bucher P. Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics.* 2003;**19**(Suppl 2):i103–i112.
5. Madera M. Profile Comparer (PRC): a program for scoring and aligning profile hidden Markov models. *Bioinformatics.* 2008;**24**:2630–2631.
6. Pachter L, Alexandersson M, Cawley S. Applications of generalized pair hidden Markov models to alignment and gene finding problems. *J. Comput. Biol.* 2002;**9**:389–399.
7. Srivastava P K, Desai D K, Nandi S, Lynn A M. HMM-ModE-improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences. *BMC Bioinform.* 2007;**8**:104.
8. S¨oding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;**21**:951–960.

9. Won K J, Hamelryck T, Prgel-Bennett A, Krogh A. An evolutionary method for learning HMM structure: prediction of protein secondary structure. *BMC Bioinform.* 2007;**8**:357.
10. Wistrand M, Sonnhammer E L. Improving profile HMM discrimination by adapting transition probabilities. *J. Mol. Biol.* 2004;**338**:847–854
11. Wistrand M, Sonnhammer E L. Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. *BMC Bioinform.* 2005;**6**:99.
12. Yoon B-J, Vaidyanathan P P. Structural alignment of RNAs using profile-csHMMs and its application to RNA homology search: Overview and new results. *IEEE Trans. Automat. Contr. (Joint Special Issue on Systems Biology with IEEE Transactions on Circuits and Systems: Part-I)* 2008;**53**:10–25.
13. Yoon B-J, Vaidyanathan PP. Context-sensitive hidden Markov models for modeling long-range dependencies in symbol sequences. *IEEE Trans. Signal Processing.* 2006;**54**:4169–4184