Pertanika J. Sci. & Technol. 26 (3): 1183 - 1198 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Novel Entropy Algorithm for State Sequence of Bakis Hidden Markov Model

Jason Chin-Tiong Chan[1]* and Hong Choon Ong[2]

[1]*Ted Rogers School of Management, Ryerson University, 350 Victoria St, Toronto, ON M5B 2K3, Canada*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Gelugor, Penang, Malaysia*

ABSTRACT

Hidden Markov model (HMM) can be categorised as an ergodic model or a left-to-right model. The categorization is subject to its state transition. An ergodic Hidden Markov model has full state transitions but a left-to-right hidden Markov model has partial state transitions. A Bakis Hidden Markov model (BHMM) is a special type of the left-to-right Hidden Markov model. State sequence for a BHMM is invisible but this research is able to track the most likelihood state sequence using Viterbi algorithm. However, while tracking the optimal state sequence for BHMM, the conventional algorithm does not provide a measure of uncertainty which is present in the solution. This issue can be overcome by the proposed novel algorithm, namely, BHMM entropy-based forward algorithm (BHMM-EFA) for computing state entropy of a BHMM. This algorithm is based on a decreasing-ladder trellis structure which provides a clear picture on how the entropy associated with the optimal state sequence is determined. Therefore, the novel algorithm requires $O(TN)$ calculations for tracking the optimal state sequence of a first-order BHMM where T is the length of the observational sequence and N is the number of hidden states.

Keywords: Bakis Hidden Markov model, entropy, forward probability, state transition, uncertainty, Viterbi Algorithm

ARTICLE INFO

Article history:
Received: 22 August 2017
Accepted: 08 February 2018

E-mail addresses:
chintiongjason.chan@ryerson.ca (Jason Chin-Tiong Chan)
hcong@usm.my (Hong Choon Ong)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

A left-to-right Hidden Markov Model (LR-HMM) is an important subclass of Hidden Markov Models for modelling time series data (Rabiner, 1989). This subclass model is a single directional structure model and hence named a left-to-right Hidden Markov Model. This subclass of model structure is useful in applications such as speech recognition (Gales & Young, 2008; Nogueries et al., 2001; Juang

& Rabiner, 1991) and DNA sequence analysis (Regad, 2008; Loytynoja & Milinkovitch, 2003). HMM is a statistical sequence model where the observational sequence is generated independently but conditioned on a hidden state Markov chain. The index of the hidden state for LR-HMM either increases or maintains the same state as time progresses. As a result, the transition matrix of LR-HMM is an upper triangular matrix with zeros on the below of the main diagonal entries, but there is at least one non-zero entry on or above the diagonal per column. State transitions between non-adjacent states are permitted in this LR-HMM, for example, a transition from state 2 to state 6. In some applications such as speech recognition (Nogueries et al., 2001), it requires more constraints on the state transition structure so that only transitions among neighbouring states are allowed. This type of special LR-HMM is called as a Bakis Hidden Markov Model (BHMM). There are two ways of state transitions for BHMM: self-state transition (i.e., from state i to state i) and one-step transitions (i.e., from state i to state i + 1). Hence, there is at least one entry for the main diagonal and the first upper diagonal is non-zero per row for the state transition matrix structure in a BHMM. The hidden state sequence of HMM can be tracked from a given observational sequence and the restored state has many applications especially when the hidden state sequence has meaningful interpretations in terms of prediction. For example, Ciriza et al. (2011) determined the optimal printing rate based on the HMM model parameter and an optimal time-out which is based on the restored states. Viterbi algorithm is the most common technique for tracking hidden state sequence (Rabiner, 1989). However, it does not measure the uncertainty present in the solution. Proakis and Salehi (2002) proposed a method for measuring the error of a single state. This method is unable to measure the error of the entire state sequence computed by Viterbi algorithm. Hernando et al. (2005) proposed a method of using entropy for measuring the uncertainty of the state sequence of a first-order HMM tracked from a single observational sequence with a length of T. The method is based on the forward recursion algorithm integrated with entropy. Mann and McCallum (2007) developed an algorithm for computing the subsequent constrained entropy of HMM which is similar to probabilistic model conditional random fields (CRF). Ilic (2011) developed an algorithm based on forward-backward recursion over the entropy semiring, namely Entropy Semiring Forward-backward algorithm (ESRFB) for a first-order HMM with a single observational sequence. ESRFB has lower memory requirement as compared with Mann and McCallum's algorithm for subsequent constrained entropy computation.

This paper is organised as follows. In section 2, a Bakis Hidden Markov model is defined and a modified forward probability variable is introduced in a decreasing-ladder trellis structure diagram. In section 3, an algorithm is proposed for computing state entropy for a Bakis Hidden Markov Model, namely BHMM entropy-based forward algorithm (BHMM-EFA) and a numerical example of computing state entropy is illustrated. We discuss future research in section 4 on this state entropy algorithm.

FIRST-ORDER DISCRETE BAKIS HIDDEN MARKOV MODEL

In this section, definitions and notations for a first-order BHMM is introduced. These are followed by the definition of the modified forward probability variable for a first-order BHMM. These forward probability variables form the basis for the proposed algorithm, namely BHMM entropy-based forward algorithm (BHMM-EFA). A trellis structure diagram is then defined and a a decreasing-ladder trellis structure diagram is introduced which provides a clear picture on how the entropy associated with the optimal state sequence of BHMM is determined.

Element of BHMM

BHMM involves two stochastic processes, namely hidden state process and observation process. The hidden state process cannot be directly observed. The observation sequence is generated by the observation process incorporated with the hidden state process. For discrete BHMM, it has to satisfy the following conditions:

The hidden state process $\{q_t\}_{t=1}^{T}$ is a first-order Bakis Markov chain that satisfies

$$P\left(q_t \middle| \{q_l\}_{l<t}\right) = P\left(q_t \middle| q_{t-1}\right) \tag{1}$$

where q_t denotes the hidden state at time t, and $q_t \in S$ where S is the finite set of hidden states. Note that if $q_{t-1}$is at the state s_i then q_i is either at the state s_i or state s_{i+1} due to self-transition or one-step transition.

The observation process $\{o_t\}_{t=1}^{T}$ is incorporated into the hidden state process according to the state probability distribution that satisfies

$$P\left(o_t \middle| \{o_l\}_{l<t}, \{q_l\}_{l\le t}\right) = P\left(o_t \middle| q_t\right) \tag{2}$$

where o_t denotes the observation at time t, and $o_t \in V$ where V is the finite set of observation symbols.

The elements for a first-order discrete BHMM are as follows:

- Number of distinct hidden states, N
- Number of distinct observed symbols, M
- Length of observed outputs, T
- Observed output sequence, $O = \{o_t, t = 1,2,\ldots,T\}$
- Hidden state sequences, $Q = \{q_t, t = 1,\ldots,T\}$
- Possible values for each state, $S = \{s_i, i = 1,2,\ldots,N\}$
- Possible symbols per observation, $V = \{v_w, w = 1,2,\ldots,M\}$
- Initial hidden state probability vector, π_i

  π_i is the probability that the model will start from state s_i

  $$\pi_i = P(q_1 = s_i)\ ,\ \sum_{i=1}^{N} \pi_i = 1, \pi_i \ge 0,\ 1 \le i \le N$$

- State transition probability matrix, $A = \{a_{ij}\}$

A is the two-dimensional state transition probability matrix, and a_{ij}, is the probability of a transition to state s_j given that it has had a transition from state s_i where $1 \le i, j \le N$.

$$a_{ij} = P(q_t = s_j | q_{t-1} = s_i), \sum_{j=i}^{i+1} a_{ij} = 1, a_{ij} \ge 0$$

- Emission output probability matrix, $B = \{b_i(v_m)\}$

B is the two-dimensional emission output probability matrix, and $b_i(v_m)$ is a probability of observing v_m in state s_i where $1 \le i \le N$.

$$b_i(v_m) = P(o_t = v_m | q_t = s_i), \quad \sum_{m=1}^{M} b_i(v_m) = 1, b_i(v_m) \ge 0$$

For a first-order discrete BHMM, parameters by using the components of $\lambda = (\pi, A, B)$ are summarised.

Note that throughout this paper, we will use the following notations.

- $q_{1:t}$ denotes $q_1, q_2, ..., q_t$
- $o_{1:t}$ denotes $o_1, o_2, ..., o_t$

Forward Probability for BHMM

Hernando et al. (2005) proposed an algorithm that incorporated the forward recursion process for computing the entropy of state sequence. Ilic (2011) also developed an algorithm based on forward-backward recursion process for state entropy computation. Both algorithms are formulated for a first-order HMM with a single observation sequence. The conventional forward probability based on the state transition structure of BHMM is modified. The BHMM has two types of state transition, which are self-state and one-step transition. These forward probabilities are computed recursively in the recursion and the termination phase. The algorithms proposed by Hernando et al. (2005) and Ilic (2011) require $\mathrm{O}(TN^2)$ calculations for computing state entropy of a generalised first-order HMM. The new algorithm performs state entropy computation that requires $\mathrm{O}(TN)$ due to its modified forward probability. The conventional forward probability for a first-order HMM is defined as follows (Rabiner, 1989):

Definition 1. The conventional forward variable $\alpha_t(i)$ for a first-order HMM is a joint probability of the partial observation sequence $o_1, o_2, \cdots, o_t$ and the hidden state s_i of at time t where $1 \le t \le T$. It can be denoted as

$$\alpha_t(i) = P(o_1, o_2, \cdots, o_t, q_t = s_i | \lambda) \quad (3)$$

From (3), $t = 1$ and $1 \le i \le N$, the initial forward variable can be expressed as

$$\begin{aligned}\alpha_1(i) &= P(o_1, q_1 = s_i|\lambda) = P(q_1 = s_i)P(o_1|q_1 = s_i)\\ &= \pi_i b_i(o_1)\end{aligned} \tag{4}$$

From (3), (4), the recursive forward variable for BHMM is obtained where $t = 2,...,T$,

$$\begin{aligned}\alpha_t(j) &= P(o_1, o_2, \ldots, o_t, q_t = s_j|\lambda)\\ &= \sum_{i=j-1}^{j} P(o_1, o_2, \ldots, o_t, q_{t-1} = s_i, q_t = s_j|\lambda)\\ &= \sum_{i=j-1}^{j} P(o_1, o_2, \ldots, o_{t-1}, q_{t-1} = s_i|\lambda)\, P(q_t = s_j|q_{t-1} = s_i)P(o_t|q_t = s_j)\\ &= \sum_{i=j-1}^{j} \alpha_{t-1}(i)a_{ij}b_j(o_t)\end{aligned} \tag{5}$$

Note that the above summation is from j-1 to j at time t due to the state transition structure of BHMM whereas the summation is from 1 to N for the conventional recursive forward probability where N is the number of the hidden states.

The recursive modified forward probability in (5) can be represented in the following form

$$\alpha_t(j) = \begin{cases} \alpha_{t-1}(j)a_{11}b_1(o_t) & \text{if} \quad j = 1 \\ \left[\alpha_{t-1}(j-1)a_{(j-1)(j)} + \alpha_{t-1}(j)a_{jj}\right]b_j(o_t) & \text{if} \quad 2 \le j \le N \end{cases} \tag{6}$$

The modified forward variable is normalised that is required as an intermediate variable in our proposed algorithm. The following is the definition for a normalised forward variable which is used for obtaining a normalising modified forward variable.

Definition 2. The normalising modified forward probability variable $\hat{\alpha}_t(i)$ in a first-order BHMM is defined as the probability of the hidden state of s_i at time given the partial observation sequence $o_1, o_2, \cdots, o_t$ where $1 \le t \le T$.

$$\hat{\alpha}_t(i) = P(q_t = s_i|o_1, o_2, \cdots, o_t) \tag{7}$$

From (4), (7), t=1 and $1 \le i \le N$, the initial normalised forward variable is obtained as

$$\begin{aligned}\hat{\alpha}_1(j) &= P(q_1 = s_j|o_1)\\ &= \frac{P(q_1 = s_j, o_1)}{P(o_1)}\\ &= \frac{\pi_j b_j(o_1)}{r_0}\end{aligned} \tag{8}$$

$$\text{where } r_0 = \sum_{k=1}^{N} \pi_k b_k(o_1) \tag{9}$$

From (5), (7), (8) and $t = 2, \cdots, T$, $1 \le j \le N$, the following is obtained

$$\begin{aligned} \hat{\alpha}_t(j) &= P(q_t = s_j | o_1, o_2, \ldots, o_t) \\ &= \frac{P(q_t = s_j, o_1, o_2, \ldots, o_t)}{P(o_1, o_2, \ldots, o_t)} \\ &= \frac{\sum_{i=j-1}^{j} \alpha_{t-1}(i) a_{ij} b_j(o_t)}{r_t} \end{aligned} \tag{10}$$

$$\text{where } r_t = \sum_{l=1}^{N} \sum_{k=l-1}^{l} \alpha_{t-1}(k) a_{kl} b_l(o_t) \tag{11}$$

Note that normalisation factors r_i ensure the modified forward probabilities sum to one.

The normalised initial and recursive forward probabilities given by (8) and (10) are used in the current algorithm for computing the recursive state entropy of a first-order BHMM. The majority of recursive entropy computations are performed in the recursion phases that results in $O(TN)$ operations. It only requires memory space of $2N$ since the memory space is independent of the length of the observational sequence. A numerical illustration is shown in section 3.2 Hernando et al. (2005) use a trellis diagram to show the structure for the recursive computation of the entropy. The time is shown on the horizontal axis. This diagram contains both states and its entropy for each time step. The Figure 1 is an example of a trellis diagram displaying the structure for the recursive computation of the entropy for a generalised first-order HMM with 4 states and a length T of observational sequence.

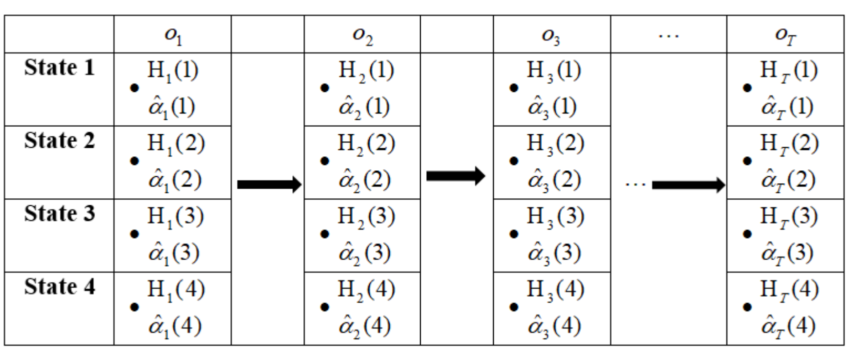

Figure 1. The evolution of a trellis structure with 4 states and a length of T observational sequence

Due to the structure of state transition for BHMM, a modified trellis structure diagram is introduced, namely a decreasing-ladder trellis structure diagram which provides a clear picture on how the entropy associated with the optimal state sequence of BHMM is determined. Figure 2 is an example of a decreasing-ladder trellis structure diagram with 4 states and 6 observations.

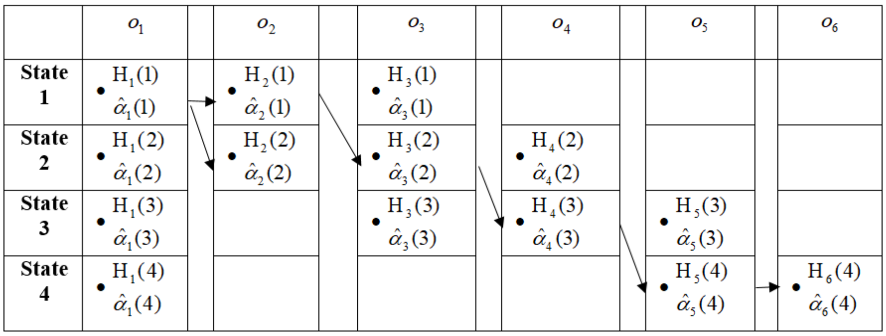

Figure 2. The evolution of a decreasing-ladder trellis structure with 4 states and 6 observations

STATE ENTROPY COMPUTATION FOR A FIRST-ORDER BHMM

The optimal state sequence can be obtained from a given BHMM model's parameters and observational sequence using the conventional Viterbi algorithm. This algorithm provides the solution along with its likelihood. This likelihood probability can be determined as follows.

$$P(q_1, q_2, \cdots, q_T | o_1, o_2, \cdots, o_T) = \frac{P(q_1, q_2, \cdots, q_T, o_1, o_2, \cdots, o_T)}{P(o_1, o_2, \cdots, o_T)}$$

Hence, this probability can be used as a measure of quality of the solution. The higher the probability of the "solution", the better it is. However, this conventional algorithm does not provide a clear process for obtaining the optimal state sequence and its likelihood probability. Alternatively, entropy is proposed for measuring the quality of the state sequence and using a decreasing-ladder trellis structure to illustrate the process of computing the possible state sequence along with the uncertainty.

The entropy H*(X)* of a random variable of *X*, is a measure of its uncertainty (Cover & Thomas, 2006). This concept for quantifying the uncertainty of the state sequence tracked by the BHMM's parameter is applied and a given single observational sequence. This entropy can be viewed as a measure of how well the parameters for generating a certain observational sequence. The entropy of the state sequence equals to 0 if there is only one state sequence that could have generated the observation sequence as there is no uncertainty in the solution. The higher this entropy, the higher the uncertainty involved in tracking the hidden state.

Entropy of a discrete random variable is defined as follows (Cover & Thomas, 2006):

Definition 3. The entropy H(X) of a discrete random variable X with a probability mass function $P(X=x)$ is defined as

$$\mathrm{H}(X) = -\sum_{x \in X} P(x) \log_2 P(x) \tag{12}$$

When the log has a base of 2, the unit of the entropy is bits and it is noted that $0 \log 0 = 0$

From (12), the entropy of the distribution for all possible state sequences is defined as follows:

$$\mathrm{H}(q_{1:T} | o_1, o_2, \cdots, o_T) = -\sum_{S} \left[P(q_{1:T} = s_{1:T} | o_1, o_2, \cdots, o_T) \log_2 P(q_{1:T} = s_{1:T} | o_1, o_2, \cdots, o_T) \right] \tag{13}$$

BHMM Entropy-based Forward Algorithm (BHMM-EFA)

For this algorithm, an intermediate variable that is state entropy, $H_t(s_j)$ is required. The state entropy, $H_t(s_j)$ can be computed recursively using the previous one that is $H_{t-1}(s_j)$ (Hernando et al., 2005).

The state entropy for a first-order BHMM is defined as follows:

Definition 4. The state entropy, $H_t(s_j)$ in a first-order BHMM, is the entropy of all the state sequences that lead to state of at time , given the observations $o_1, o_2, \cdots, o_t$. It can be denoted as

$$\mathrm{H}_t(s_j) = \mathrm{H}(q_{1:t-1} | q_t = s_j, o_{1:t}) \tag{14}$$

From (14) and t=1, the initial state entropy variable is expressed as

$$\mathrm{H}_1(s_j) = 0 \tag{15}$$

From (14), and (15), the recursion on the entropies for $t = 2, \ldots, T$, $1 \le i, j \le N$, and $i = j-1, j$ is obtained as below

$$\begin{aligned} \mathrm{H}_t(s_j) &= \mathrm{H}(q_{1:t-1} | q_t = s_j, o_{1:t}) \\ &= \mathrm{H}(q_{1:t-2}, q_{t-1} | q_t = s_j, o_{1:t}) \\ &= \mathrm{H}(q_{t-1} | q_t = s_j, o_{1:t}) + \mathrm{H}(q_{1:t-2} | q_{t-1}, q_t = s_j, o_{1:t}) \end{aligned} \tag{16}$$

where

$$\mathrm{H}(q_{t-1} | q_t = s_j, o_{1:t}) = -\sum_{i=j-1}^{j} \left[P(q_{t-1} = s_i | q_t = s_j, o_{1:t}) \log_2 \left(P(q_{t-1} = s_i | q_t = s_j, o_{1:t}) \right) \right]$$

and

$$\mathrm{H}\left(q_{1:t-2} \middle| q_{t-1}, q_t = s_j, o_{1:t}\right) = \sum_{i=j-1}^{j}\left[P\left(q_{t-1} = s_i \middle| q_t = s_j, o_{1:t}\right)\mathrm{H}\left(q_{1:t-2} \middle| q_{t-1} = s_i, q_t = s_j, o_{1:t}\right)\right]$$

$$= \sum_{i=j-1}^{j}\left[P\left(q_{t-1} = s_i \middle| q_t = s_j, o_{1:t}\right)\mathrm{H}_{t-1}(s_i)\right]$$

The auxiliary probability $P\left(q_{t-1} = s_i \middle| q_t = s_j, o_{1:t}\right)$ where $i = j-1, j$ is required for our algorithm. It can be computed as follows:

$$\begin{aligned}
&P\left(q_{t-1} = s_i \middle| q_t = s_j, o_{1:t}\right) = P\left(q_{t-1} = s_i \middle| q_t = s_j, o_t, o_{1:t-1}\right)\\
&= \frac{P\left(q_{t-1} = s_i, q_{\mathrm{t}} = s_j, o_{1:t-1}, o_t\right)}{P\left(q_{\mathrm{t}} = s_j, o_{1:t-1}, o_t\right)}\\
&= \frac{P\left(q_{\mathrm{t}} = s_j, o_t \middle| q_{t-1} = s_i, o_{1:t-1}\right)P\left(q_{\mathrm{t-1}} = s_i \middle| o_{1:t-1}\right)}{P\left(q_{\mathrm{t}} = s_j, o_t \middle| o_{1:t-1}\right)}\\
&= \frac{P\left(o_t \middle| q_t = s_j\right)P\left(q_t = s_j \middle| q_{t-1} = s_i\right)P\left(q_{t-1} = s_i \middle| o_{1:t-1}\right)}{P\left(o_t \middle| q_t = s_j\right)P\left(q_t = s_j \middle| o_{1:t-1}\right)} \qquad (17)\\
&= \frac{P\left(q_t = s_j \middle| q_{t-1} = s_i\right)P\left(q_{t-1} = s_i \middle| o_{1:t-1}\right)}{\sum_{k=j-1}^{j} P\left(q_t = s_j \middle| q_{t-1} = s_k\right)P\left(q_{t-1} = s_k \middle| o_{1:t-1}\right)}\\
&= \frac{a_{ij}\hat{\alpha}_{t-1}(i)}{\sum_{k=j-1}^{j} a_{kj}\hat{\alpha}_{t-1}(k)}
\end{aligned}$$

For the final process, $\mathrm{H}\left(q_{1:T} \middle| o_{1:T}\right)$ is computed which can be expanded as follows:

$$\begin{aligned}
\mathrm{H}\left(q_{1:T} \middle| o_{1:T}\right) &= \mathrm{H}\left(q_{1:T-1}, q_T \middle| o_{1:T}\right) = \mathrm{H}\left(q_{1:T-1} \middle| q_T, o_{1:T}\right) + \mathrm{H}\left(q_T \middle| o_{1:T}\right)\\
&= \sum_{i=1}^{N} \mathrm{H}_T(s_i)\hat{\alpha}_T(i) - \sum_{i=1}^{N} \hat{\alpha}_T(i)\log_2\left(\hat{\alpha}_T(i)\right)
\end{aligned} \qquad (18)$$

The basic entropy concept in (12) and the following basic properties of BHMM are used for proving lemma 1. According to a first-order BHMM, state q_{t-r}, $r \geq 2$ and q_t are statistically independent given q_{t-1}. The same applies to q_{t-r}, $r \geq 2$ and o_t are statistically independent given q_{t-1}.

The following proof is due to Hernando et. al. (2005).

Lemma 1: For a first-order BHMM, the entropy of the state sequence up to time t - 2, given the states at time t - 1 and the observations up to time t - 1, is conditionally independent on the state and observation at time t

$$\mathrm{H}_{t\text{-}1}(s_i) = \mathrm{H}\left(q_{1:t-2} \middle| q_{t-1} = s_i, q_t = s_j, o_{1:t}\right)$$

Proof:

$$\mathrm{H}\left(q_{1:t-2} \middle| q_{t-1}=s_i, q_t=s_j, o_{1:t}\right)$$
$$=\mathrm{H}\left(q_{1:t-2} \middle| q_{t-1}=s_i, o_{1:t-1}, q_t=s_j, o_t\right)$$
$$=\mathrm{H}\left(q_{1:t-2} \middle| q_{t-1}=s_i, o_{1:t-1}\right)$$
$$=\mathrm{H}_{t\text{-}1}\left(s_i\right)$$

The entropy algorithm for BHMM is based on normalised forward recursion probability variable, state entropy recursion variable and auxiliary probability.

From (8), (10), (15), (16), (17) and (18), the algorithm of entropy computation for a first-order BHMM is constructed as below:

1. Initialisation: for $1 \le j \le N$,

$$\mathrm{H}_1\left(s_j\right)=0$$

$$\hat{\alpha}_1(j)=\frac{\pi(j)b_1(o_1)}{\sum_{k=1}^{N}\pi(k)b_k(o_1)}$$

2. Recursion:

For $t=2:T$

For $j=1:N-1$

If $\hat{\alpha}_{t-1}(j)>0$

$w=j$

For $j=w:w+1$

$$\hat{\alpha}_t(j)=\frac{\sum_{i=j-1}^{j}\hat{\alpha}_{t-1}(i)a_{ij}b_j(o_t)}{\sum_{l=1}^{N}\sum_{k=l-1}^{l}\hat{\alpha}_{t-1}(k)a_{kl}b_l(o_t)}$$

For $i=j-1:j$

$$P\left(q_{t-1}=s_i \middle| q_t=s_j, o_{1:t}\right)=\frac{a_{ij}\hat{\alpha}_{t-1}(i)}{\sum_{k=j-1}^{j}a_{kj}\hat{\alpha}_{t-1}(k)}$$

End

$$\mathrm{H}_t\left(s_j\right)=\sum_{i=j-1}^{j}\mathrm{H}_{t\text{-}1}\left(s_i\right)P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)$$
$$-\sum_{i=j-1}^{j}P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)\log_2 P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)$$

End

End if

End

If $\hat{\alpha}_{t-1}(N)>0$

$j=N$

$$\hat{\alpha}_t\left(j\right)=\frac{\sum_{i=j-1}^{j}\hat{\alpha}_{t-1}\left(i\right)a_{ij}b_j\left(o_t\right)}{\sum_{l=1}^{N}\sum_{k=l-1}^{l}\hat{\alpha}_{t-1}\left(k\right)a_{kl}b_l\left(o_t\right)}$$

For $i=N-1:N$

$$P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)=\frac{a_{ij}\hat{\alpha}_{t-1}\left(i\right)}{\sum_{k=j-1}^{j}a_{kj}\hat{\alpha}_{t-1}\left(k\right)}$$

End

$$\mathrm{H}_t\left(s_j\right)=\sum_{i=j-1}^{j}\mathrm{H}_{t\text{-}1}\left(s_i\right)P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)$$
$$-\sum_{i=j-1}^{j}P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)\log_2 P\left(q_{t-1}=s_i\middle|q_t=s_j,o_{1:t}\right)$$

End if

End

3. Termination

$$\mathrm{H}\left(q_{1:T}\middle|o_{1:T}\right)=\sum_{i=1}^{N}\mathrm{H}_T\left(s_i\right)\hat{\alpha}_T\left(i\right)-\sum_{i=1}^{N}\hat{\alpha}_T\left(i\right)\log_2\left(\hat{\alpha}_T\left(i\right)\right)$$

Illustration

A BHMM is considered with the following model parameters $\pi = (\lambda, A, B)$ where π is the initial transition probability vector, A is the state transition probability matrix and B is the emission probability matrix.

$$\pi = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix},\ A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The algorithm for computing the state entropy is applied based on the observational sequence $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$. The result is shown as follows.

	$o_1 = 1$	$o_2 = 2$	$o_3 = 2$	$o_4 = 3$	$o_5 = 5$	$o_6 = 5$
State 1	$H_1(1) = 0$ $\hat{\alpha}_1(1) = 1$	$H_2(1) = 0$ $\hat{\alpha}_2(1) = 0$				
State 2	$H_1(2) = 0$ $\hat{\alpha}_1(2) = 0$	$H_2(2) = 0$ $\hat{\alpha}_2(2) = 1$	$H_3(2) = 0$ $\hat{\alpha}_3(2) = 1$	$H_4(2) = 0$ $\hat{\alpha}_4(2) = 0.5$	$H_5(2) = 0$ $\hat{\alpha}_5(2) = 0$	
State 3	$H_1(3) = 0$ $\hat{\alpha}_1(3) = 0$		$H_3(3) = 0$ $\hat{\alpha}_3(3) = 0$	$H_4(3) = 0$ $\hat{\alpha}_4(3) = 0.5$	$H_5(3) = 1$ $\hat{\alpha}_5(3) = 0$	
State 4	$H_1(4) = 0$ $\hat{\alpha}_1(4) = 0$				$H_5(4) = 0$ $\hat{\alpha}_5(4) = 1$	$H_6(4) = 0$ $\hat{\alpha}_6(4) = 1$
Entropies	**0**	**0**	**0**	**1**	**0**	**0**

Figure 3. The evolution of the decreasing-ladder trellis structure with the observation sequence $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$

The total entropy after each time step is displayed at the bottom of Figure 3. For example, after receiving the fourth observation, i.e. $o_{1:4} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3)$, it has produced two possible state sequences which are $q_{1:4} = (q_1 = state1, q_2 = state2, q_3 = state2, q_4 = state2)$ and $q_{1:4} = (q_1 = state1, q_2 = state2, q_3 = state2, q_4 = state3)$ as shown by the arrows. Each possible state sequence has a probability of 0.5 and hence the entropy is 1 bit. After receiving the fourth observation, intermediate entropy and normalized forward probability for state 2 and state 3 are computed whereas both values for state 1 and state 4 are zeros due to the structure of the state transition. For the fifth observation, the intermediate entropy and normalized forward probability for state 2, state 3 and state 4 are computed whereas both values for state 1 are zeros due to the structure of the state transition. The following are the required computations after receiving the fourth observation, i.e. $o_{1:4} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3)$ in order to compute the total entropy at t = 4. The model has produced only one possible state sequence that is $q_{1:3} = (q_1 = state1, q_2 = state2, q_3 = state2)$ with a probability of 1 after receiving the third observation. The total entropy is 0 at which $t = 3$ indicates that there is no uncertainty. Due to the structure of the state transition, at time $t = 4$, only two state transitions occurred that are the transition from state 2 at $t = 3$ to state 2 at $t = 4$ and state 2 at $t = 3$ to state 3 at $t = 4$.

From (10) and $j = 2$, we obtain

$$\hat{\alpha}_4(2) = \frac{\hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4)}{\hat{\alpha}_3(1)a_{11}b_1(o_4) + \hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4) + \hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4) + \hat{\alpha}_3(3)a_{34}b_4(o_4) + \hat{\alpha}_3(4)a_{44}b_4(o_4)}$$

$$= \frac{0(0.5)(0.5) + 1(0.5)(0.5)}{0(0.5)(0) + 0(0.5)(0.5) + 1(0.5)(0.5) + 1(0.5)(0.5) + 0(0.5)(0.5) + 0(0.5)(0) + 0(1)(0)} = 0.5$$

From (10) and $j = 3$, we obtain

$$\hat{\alpha}_4(3) = \frac{\hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4)}{\hat{\alpha}_3(1)a_{11}b_1(o_4) + \hat{\alpha}_3(1)a_{12}b_2(o_4) + \hat{\alpha}_3(2)a_{22}b_2(o_4) + \hat{\alpha}_3(2)a_{23}b_3(o_4) + \hat{\alpha}_3(3)a_{33}b_3(o_4) + \hat{\alpha}_3(3)a_{34}b_4(o_4) + \hat{\alpha}_3(4)a_{44}b_4(o_4)}$$

$$= \frac{1(0.5)(0.5) + 0(0.5)(0.5)}{0(0.5)(0) + 0(0.5)(0.5) + 1(0.5)(0.5) + 1(0.5)(0.5) + 0(0.5)(0.5) + 0(0.5)(0) + 0(1)(0)} = 0.5$$

From (17) and $i = 1, j = 2$, we obtain

$$P(q_3 = s_1 | q_4 = s_2, o_{1:4}) = \frac{a_{12}\hat{\alpha}_3(1)}{a_{12}\hat{\alpha}_3(1) + a_{22}\hat{\alpha}_3(2)} = \frac{0.5(0)}{0.5(0) + 0.5(1)} = 0,$$

From (17) and $i = 2, j = 2$, we obtain

$$P(q_3 = s_2 | q_4 = s_2, o_{1:4}) = \frac{a_{22}\hat{\alpha}_3(2)}{a_{12}\hat{\alpha}_3(1) + a_{22}\hat{\alpha}_3(2)} = \frac{0.5(1)}{0.5(0) + 0.5(1)} = 1,$$

From (17) and $i = 2, j = 3$, the following is obtained

$$P(q_3 = s_2 | q_4 = s_3, o_{1:4}) = \frac{a_{23}\hat{\alpha}_3(2)}{a_{23}\hat{\alpha}_3(2) + a_{33}\hat{\alpha}_3(3)} = \frac{0.5(1)}{0.5(1) + 0.5(0)} = 1,$$

From (17) and $i = 3, j = 3$ we obtain

$$P(q_3 = s_3 | q_4 = s_3, o_{1:4}) = \frac{a_{33}\hat{\alpha}_3(3)}{a_{23}\hat{\alpha}_3(2) + a_{33}\hat{\alpha}_3(3)} = \frac{0.5(0)}{0.5(1) + 0.5(0)} = 0,$$

From (16), $t = 4$ and $s_j = 2$, the following is obtained

$$\mathrm{H}_4(2) = \mathrm{H}_3(1)P(q_3 = s_1 | q_4 = s_2, o_{1:4}) + \mathrm{H}_3(2)P(q_3 = s_2 | q_4 = s_2, o_{1:4}) -$$

$$P(q_3 = s_1 | q_4 = s_2, o_{1:4})\log P(q_3 = s_1 | q_4 = s_2, o_{1:4}) - P(q_3 = s_2 | q_4 = s_2, o_{1:4})\log P(q_3 = s_2 | q_4 = s_2, o_{1:4})$$

$$= 0(0) + 0(1) - 0\log 0 - 1\log 1 = 0$$

From (16), and $t = 4$ and $s_j = 3$, we obtain

$$\mathrm{H}_4(3) = \mathrm{H}_3(2)P(q_3 = s_2 | q_4 = s_3, o_{1:4}) + \mathrm{H}_3(3)P(q_3 = s_3 | q_4 = s_3, o_{1:4}) -$$

$$P(q_3 = s_2 | q_4 = s_3, o_{1:4}) \log P(q_3 = s_2 | q_4 = s_3, o_{1:4}) - P(q_3 = s_3 | q_4 = s_3, o_{1:4}) \log P(q_3 = s_3 | q_4 = s_3, o_{1:4})$$

$$= 0(1) + 0(0) - 1\log 1 - 0 \log 0 = 0$$

From (18) and $t = 4$, the total entropy is obtained

$$\mathrm{H}(q_{1:4} | o_{1:4}) = \mathrm{H}_4(s_1)\hat{\alpha}_4(1) + \mathrm{H}_4(s_2)\hat{\alpha}_4(2) + \mathrm{H}_4(s_3)\hat{\alpha}_4(3) + \mathrm{H}_4(s_4)\hat{\alpha}_4(4) - \hat{\alpha}_4(1)\log_2(\hat{\alpha}_4(1)) -$$

$$\hat{\alpha}_4(2)\log_2(\hat{\alpha}_4(2)) - \hat{\alpha}_4(3)\log_2(\hat{\alpha}_4(3)) - \hat{\alpha}_4(4)\log_2(\hat{\alpha}_4(4))$$

$$= 0(0) + 0(0.5) + 0(0.5) + 0(0) - 0\log_2 0 - 0.5\log_2 0.5 - 0.5\log_2 0.5 - 0\log_2 0 = 1 \text{ bit.}$$

After receiving the sixth observation, i.e. $o_{1:6} = (o_1 = 1, o_2 = 2, o_3 = 2, o_4 = 3, o_5 = 5, o_6 = 5)$, the model has produced only one possible state sequence that is $q_{1:6} = (q_1 = state1, q_2 = state2, q_3 = state2, q_4 = state3, q_5 = state4, q_6 = state4)$ with a probability of 1 and hence the total entropy is 0 which indicates that there is no uncertainty.

CONCLUSION AND FUTURE WORK

An algorithm for computing the state entropy for a first-order BHMM was introduced. This algorithm needs to run with Viterbi algorithm in tracking the state sequence as well as the entropy of the distribution of the state sequence. A decreasing-ladder trellis structure is introduced in this paper. The algorithm can be represented in this structure diagram shows a clear process of obtaining the uncertainty present in state sequence. This algorithm requires $\mathrm{O}(TN)$ calculations and performs the state computation linearly with the length of observational sequence and the number of hidden states of BHMM. This research can be also extended for a discrete as well as a continuous high-order BHMM. For any generalised high-order BHMM, the historical state information is well explored for predicting the next state and these models are widely used in speech recognition.

REFERENCES

Ciriza, V., Donini, L., Durand, J., & Girard, S. (2011). *Optimal timeouts for power management under renewal or hidden Markov processes for requests.* Technical Report. Retrieved from http://hal.inria.fr/hal-00412509/en.

Cover, T. M., & Thomas. J. A. (2006). *Elements of Information Theory* (2nd Ed.). New York: Wiley.

Gales, M., & Young, S. (2008). The application of hidden Markov Models in speech recognition. *Foundations and Trends in Signal Processing, 1*(3), 195–304.

Hernando, D., Crespi, V., & Cybenko, G. (2005). Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE Transactions Information Theory, 51*(7), 2681 – 2685.

Ilic, V. M. (2011). *Entropy semiring forward-backward algorithm for HMM entropy computation.* Retrieved from ArXiv:1108.0347.

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics, 33*(3), 251–272.

Löytynoja, A., & Milinkovitch, M. C. (2003). A Hidden Markov model for progressive multiple alignment. *Bioinformatics, 19*(12), 1505–1513.

Mann, G. S., & McCallum, A. (2007). Efficient computation of entropy gradient for semi-supervised conditional random fields. In *Human Language Technologies 2007:* The *Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume. S*hort Papers on XX, NAACL '07 (pp. 109–112). Morristown, NJ, USA. Association for Computational Linguistics.

Nogueiras, A., Moreno, A., Bonafonte, A., & Marino, J. (2001). Speech emotion recognition using hidden markov models. In *Eurospeech 2001, Poster Proceedings* (pp. 2679-2682). Scandinavia.

Proakis, J. G., &Salehi, M. (2002). *Communications System Engineering.* Upper Saddler River, New Jersey: Prentice-Hall.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* (pp. 257–286). IEEE.

Regad, L., Guyon, F., Maupetit, J., Tufféry, P., & Camproux, A. C. (2008). A Hidden Markov Model applied to the protein 3D structure analysis. Computational Statistics and Data Analysis, 52(6), 3198–3207.