Hidden Markov Model

Li zegang

Xie yufeng

Zhu jiahao

Outline

- Markov Chain and Hidden Markov Model(HMM)
- Three Basic Problems for HMM
- Likelihood Computation: The Forward Algorithm
- Decoding: The Viterbi Algorithm
- HMM Training: The Forward-Backward Algorithm
- Summary

Markov Chain and Hidden Markov Model

Markov assumption:

$$P(q_t = S_j | q_{t-1} = S_i, q_{t-1} = S_k, \dots) = P(q_t = S_j | q_{t-1} = S_i) = a_{ij}$$

Formally, a Markov chain is specified by the following components:

$S = S_1 S_2 \dots S_N$ a set of N states

$A = a_{11} a_{12} \dots a_{N1} \dots a_{NN}$ a transition probability matrix A, each a_{ij} representing the probability of moving from state i to state j, s.t. $\sum_{j=1}^{N} a_{ij} = 1, \forall i$

$\pi = \pi_1, \pi_2, \dots, \pi_N$an initial probability distribution over states. π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j = 0$,meaning that they cannot be initial states. Also $\sum_{i=1}^{N} \pi_i = 1, \forall i$

Markov Chain and Hidden Markov Model

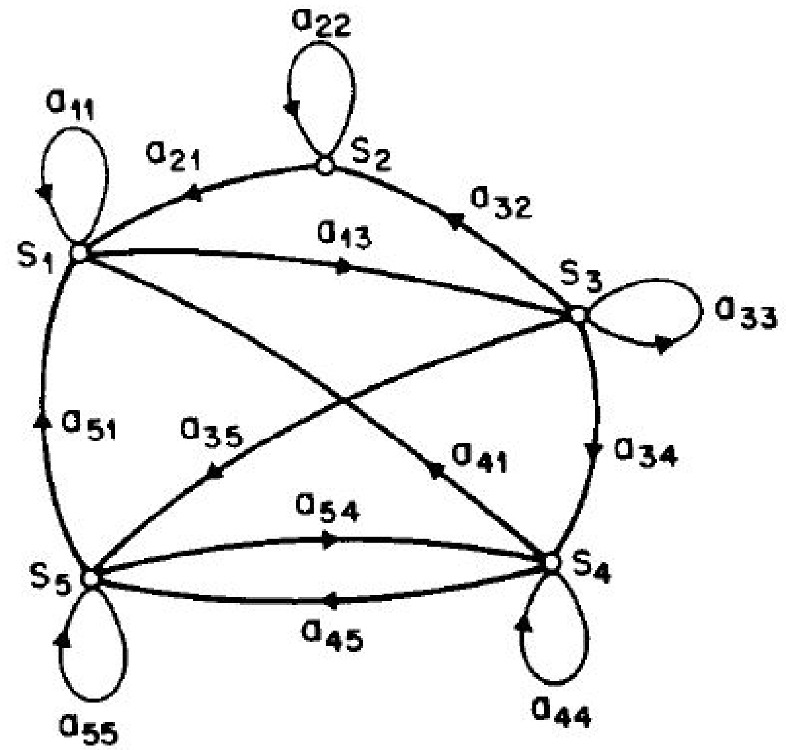
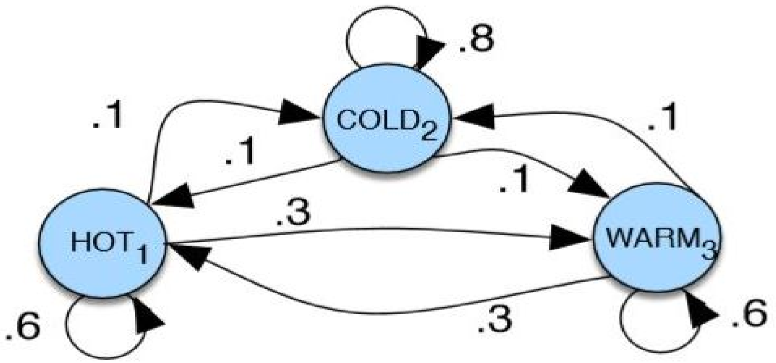


Fig. 1. A Markov chain with 5 states (labeled S_1 to S_5) with selected state transitions.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}$$

Markov Chain and Hidden Markov Model



$$A_1 = \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.1 & 0.8 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$$

Fig.2 A Markov chain for weather, showing states and transitions.

Use the sample probabilities in Fig.2 (with π = [0.1,0.7,0.2]) to compute the probability of each of the following sequences:

(1) hot hot hot hot $P(q_1 = S_1, q_2 = S_1, q_3 = S_1, q_4 = S_1)$

$$= P(q_1 = S_1)P(q_2 = S_1 | q_1 = S_1)P(q_3 = S_1 | q_1 = S_1, q_2 = S_1)P(q_4 = S_1 | q_1 = S_1, q_2 = S_1, q_3 = S_1)$$

$$=P(q_1 = S_1)P(q_2 = S_1 | q_1 = S_1)P(q_3 = S_1 | q_2 = S_1)P(q_4 = S_1 | q_3 = S_1)$$

$$= \pi_1 a_{11} a_{11} a_{11} = 0.1 * 0.6 * 0.6 * 0.6 = 0.0216$$

(2) cold hot cold hot $P(q_1 = S_2, q_2 = S_1, q_3 = S_2, q_4 = S_1) = 0.0007$

Markov Chain and Hidden Markov Model

A Markov chain is useful when we need to compute a probability for a sequence of observable events. In many cases, however, the events we are interested in are hidden: we don't observe them directly.

A hidden Markov model (HMM) allows us to talk about both observed events Markov model and hidden events that we think of as causal factors in our probabilistic model.

Markov Chain and Hidden Markov Model

Markov assumption:

$$P\left(q_t = S_j \middle| q_{t-1} = S_i, q_{t-1} = S_k, \dots\right) = P\left(q_t = S_j \middle| q_{t-1} = S_i\right) = a_{ij}$$

Output independence:

$$P\left(O_i \middle| q_1 = S_1, \dots, q_i = S_i, \dots, q_t = S_N, O_1, \dots, O_j, \dots O_T\right) = P(O_i | q_i = S_i)$$

Markov Chain and Hidden Markov Model

An HMM is specified by the following components:

$S = S_1 S_2 \dots S_N$ a set of N states

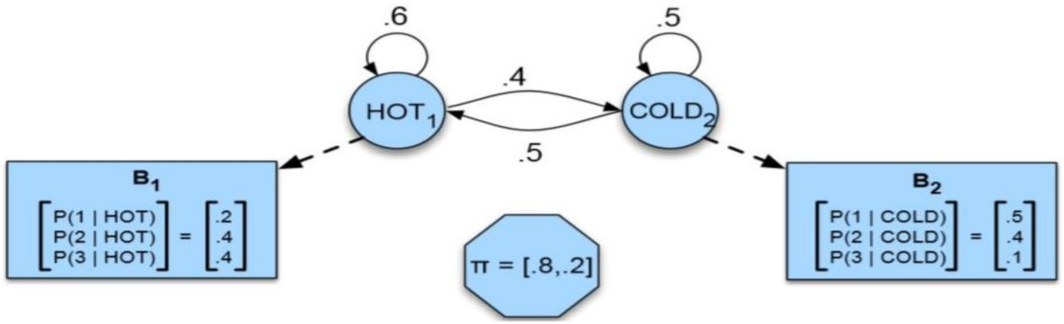
$A = a_{11} a_{12} \dots a_{N1} \dots a_{NN}$ a transition probability matrix A, each a_{ij} representing the probability of moving from state i to state j, s.t. $\sum_{j=1}^{N} a_{ij} = 1, \forall i$

$O = O_1 O_2 \dots O_T$ a sequence of T observations, each one drawn from a vocabulary $V = V_1 V_2 \dots V_V$

$B = b_i(O_t) = P(O_t | q_i = S_i)$ a sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation O_t being generated from a state i

$\pi = \pi_1, \pi_2, \dots, \pi_N$an initial probability distribution over states. π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j = 0$,meaning that they cannot be initial states. Also $\sum_{i=1}^{N} \pi_i = 1, \forall i$

Markov Chain and Hidden Markov Model



$$A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

Fig.3 A hidden Markov model for relating numbers of ice creams eaten by Jason(the observations) to the weather (H or C, the hidden variables).

Use the sample probabilities in Fig.3 (with π = [0.8,0.2]) to compute the following probability:

(1)$P(3\ 1\ 3,\ hot\ hot\ cold) = P(3\ 1\ 3|hot\ hot\ cold) * P(hot\ hot\ cold) = P(3|hot) * P(1|hot) * P(3|cold) * P(hot) * P(hot|hot) * P(cold|hot) = 0.4 * 0.2 * 0.1 * 0.8 * 0.6 * 0.4 = 0.001536$

In general, this is:

$$P(O,Q) = P(O|Q) * P(Q) = \prod_{i=1}^{T} P(O_i|q_i) * \prod_{i=1}^{T} P(q_i|q_{i-1})$$

$$P(O) = \sum_{Q} P(O,Q)$$

Three Basic Problems for HMM

- Problem1(Likelihood): Given an HMM $\lambda = (A,B)$ and an observation sequence O, determine the likelihood $P(O|\lambda)$.
- Problem2(Decoding): Given an observation sequence O and an HMM $\lambda = (A,B)$, discover the best hidden state sequence Q.
- Problem3(Learning): Given an observation sequence O and the set of states in the HMM, learn the HMM parameters A and B.

Three Basic Problems for HMM

- Problem1(Likelihood): Given an HMM $\lambda = (A,B)$ and an observation sequence O, determine the likelihood $P(O|\lambda)$.
- Problem2(Decoding): Given an observation sequence O and an HMM $\lambda = (A,B)$, discover the best hidden state sequence Q.
- Problem3(Learning): Given an observation sequence O and the set of states in the HMM, learn the HMM parameters A and B.

Likelihood Computation: The Forward Algorithm

For an HMM with N hidden states and an observation sequence of T observations, there are N^T possible hidden sequences.

The forward algorithm is an efficient algorithm with $O(N^2T)$ complexity. The forward algorithm is a kind forward algorithm of dynamic programming algorithm, that is, an algorithm that uses a table to store intermediate values as it builds up the probability of the observation sequence.

Likelihood Computation: The Forward Algorithm

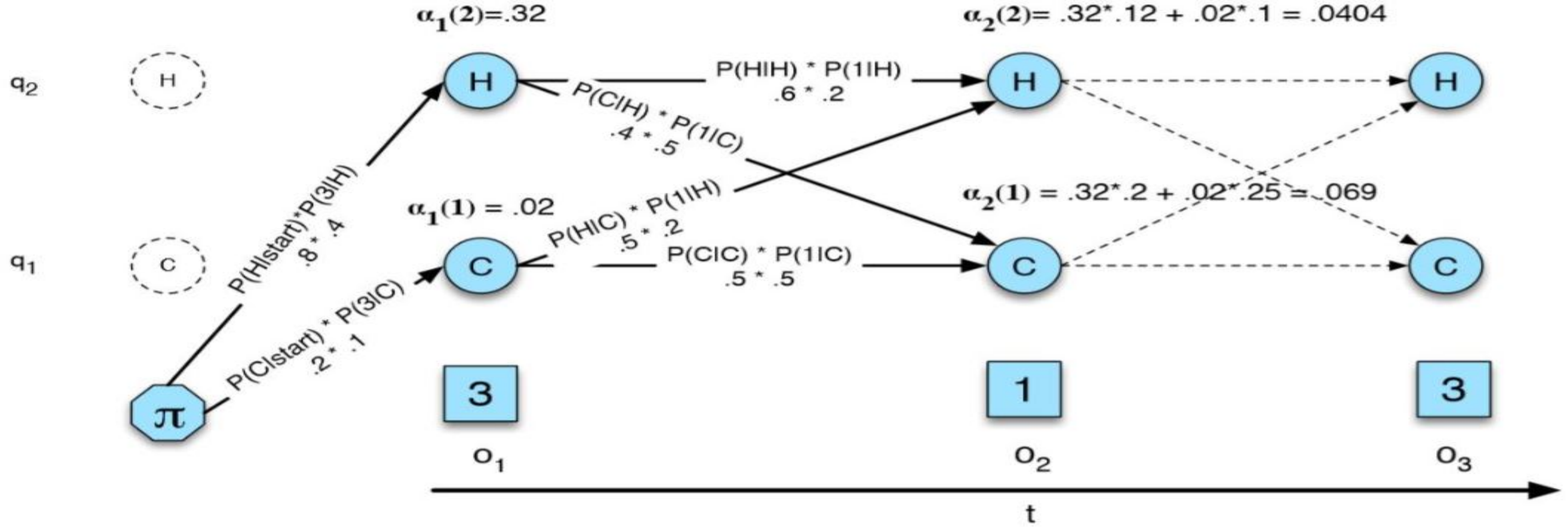


Fig.5 The forward trellis for computing the total observation likelihood for the ice-cream events 3 1 3. Hidden states are in circles, observations in squares. The figure shows the computation of $\alpha_t(S_j)$ for two states at two time steps.

$$\alpha_t(S_j) = P(O_1, \ldots, O_t, q_t = S_j)$$

Likelihood Computation: The Forward Algorithm

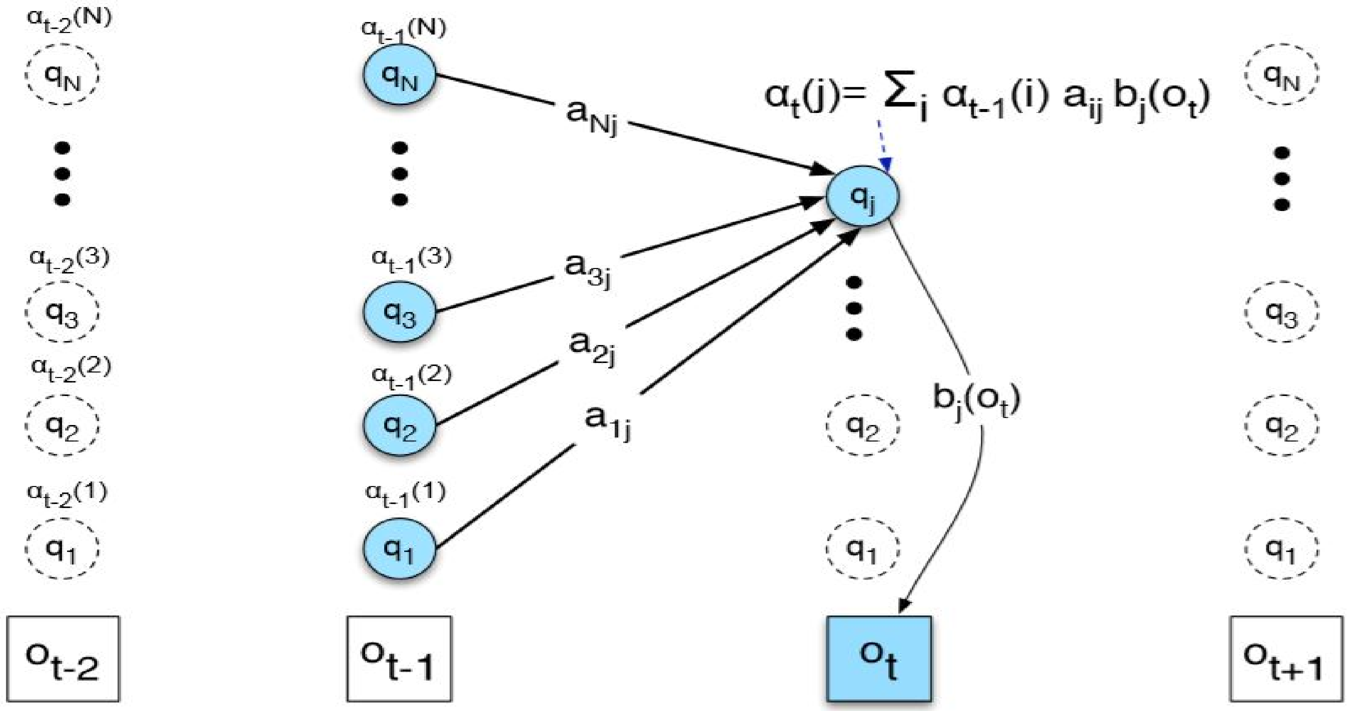


Fig.6 Visualizing the computation of a single element $\alpha_t(S_j)$ in the trellis by summing all the previous values α_{t-1}, weighted by their transition probabilities a, and multiplying by the observation probability $b_i(O_t)$. For many applications of HMMs, many of the transition probabilities are 0, so not all previous states will contribute to the forward probability of the current state. Hidden states are in circles, observations in squares. Shaded nodes are included in the probability computation for $\alpha_t(S_j)$.

Likelihood Computation: The Forward Algorithm

- Initialization

$$\alpha_1(S_j) = \pi_j b_j(O_1)$$

- Recursion

$$\alpha_t(S_j) = \sum_{i=1}^{N} \alpha_{t-1}(S_i) a_{ij} b_j(O_t)$$

- Termination

$$P(O) = \sum_{i=1}^{N} \alpha_T(S_i)$$