MACHINE INTELLIGENCE

UNIT - 3

Hidden Markov Models

feedback/corrections: vibha@pesu.pes.edu

VIBHA MASTI

1. Markov Models

Stochastic Process

- Collection of RVs where the RVs take values from a common state space
- Eg: toss a coin N times where SS = {H, T}

IID Assumption

- Independent Identically Distributed

Markov Assumption

- Probability of observing the next state only depends on current state

$$P(q_n = a | q_1 q_2 \dots q_{n-1}) = P(q_n = a | q_{n-1})$$

- Markov process of first order
- For arbitrary order k

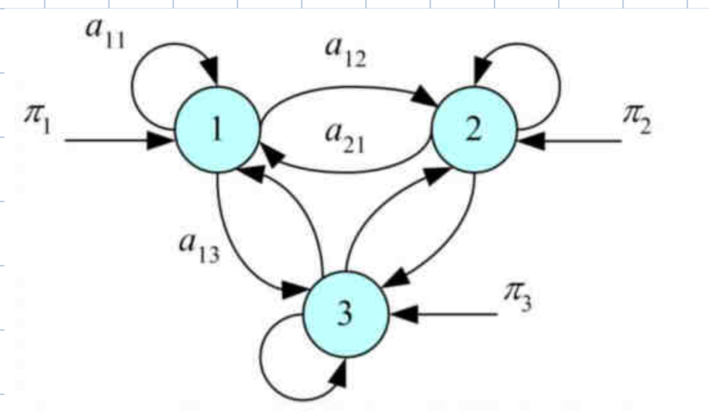
$$P(q_n = a | q_1 q_2 \dots q_{n-1}) = P(q_n = a | q_{n-1} q_{n-2} \dots q_{n-k})$$

Markov Chain

- First order Markov Process with no hidden states

Components

* set of states Q_N
* initial probabilities π_N
* transition matrix $A_{N \times N}$



Characterised as $\lambda(\pi, A)$

Meaning of Components

- π_j : initial distribution — probability that first state emitted by chain is j

$$\pi_j = P(q_1 = j)$$

- a_{ij} : transition probability — probability that chain moves from state i to state j at the time t to time $t+1$

$$a_{ij} = P(q_{t+1} = j | q_t = i)$$

Markov Chain: State Sequence Likelihood

Given :

(i) Markov Chain $\lambda(\pi, A)$
(ii) Observed state sequence Q

Solution:

Let $Q = \{q_0,\ q_1,\ \ldots\ ,\ q_m\}$

$$P(Q) = P(q_0) \times P(q_1|q_0) \times P(q_2|q_1) \times \ldots \times P(q_m|q_{m-1})$$

$P(q_0)$ can be found from π

$$P(q_0) = \pi_{q_0}$$

$P(q_i|q_j)$ can be found from A

$$P(q_i|q_j) = a_{ji}$$

Finding Parameters π and A

Given: several random walks/ state sequences
set of possible states Q

Solution:

(i) Finding π

For every state $q_i \in Q$

$$\pi_{q_i} = \frac{Number\ of\ sequences\ that\ start\ with\ q_i}{Total\ number\ of\ sequences\ given}$$

Sum of all values in π must equal 1

(ii) Finding A

For every $q_i, q_j \in Q$

$$a_{ij} = \frac{Number\ of\ transitions\ from\ state\ q_i\ to\ state\ q_j}{Number\ of\ transitions\ starting\ at\ state\ q_i}$$

Row sum must be 1

Q: Given $Q = (S, C, W)$, the following set of sequences, find HMM parameters A, π.

1. S C W S C
2. S W C S S
3. W C S W C
4. C S C W C
5. C W S S W
6. W C S C C
7. W C C W S
8. W C W C S
9. S W C W S
10. W W C C S
11. C C S W S
12. W C S W S
13. W S W C S
14. W S W S W

$$\pi(S, C, W) = \left(\frac{3}{14}, \frac{3}{14}, \frac{8}{14}\right)$$

$$A = \begin{bmatrix} \frac{2}{15} & \frac{4}{15} & \frac{9}{15} \\ \frac{9}{19} & \frac{4}{19} & \frac{6}{19} \\ \frac{9}{22} & \frac{12}{22} & \frac{1}{22} \end{bmatrix}$$

2. Hidden Markov Models

- Internal states $q_1, q_2, \ldots, q_n$ hidden
- Only observations visible
- Whenever HMM reaches a state, it emits an observation based on the emission probability that depends on the state

Components

* set of hidden states Q_N
* set of observed states O_M (alphabet)
* initial probabilities π_N
* transition matrix $A_{N \times N}$
* emission matrix $B_{N \times M}$

Characterised as $\lambda(\pi, A, B)$

Meaning of Components

- π_j : initial distribution — probability that first state emitted by chain is j

$$\pi_j = P(q_1 = j)$$

- a_{ij} : transition probability — probability that chain moves from state i to state j at the time t to time $t+1$

$$a_{ij} = P(q_{t+1} = j | q_t = i)$$

- $b_j(O_t)$: emission probability — probability that observation at time t (O_t) is emitted by state j

$$b_j(O_t) = P(o_t = O_t | q_t = j)$$

Fundamental Assumptions

(1) Markov Assumption

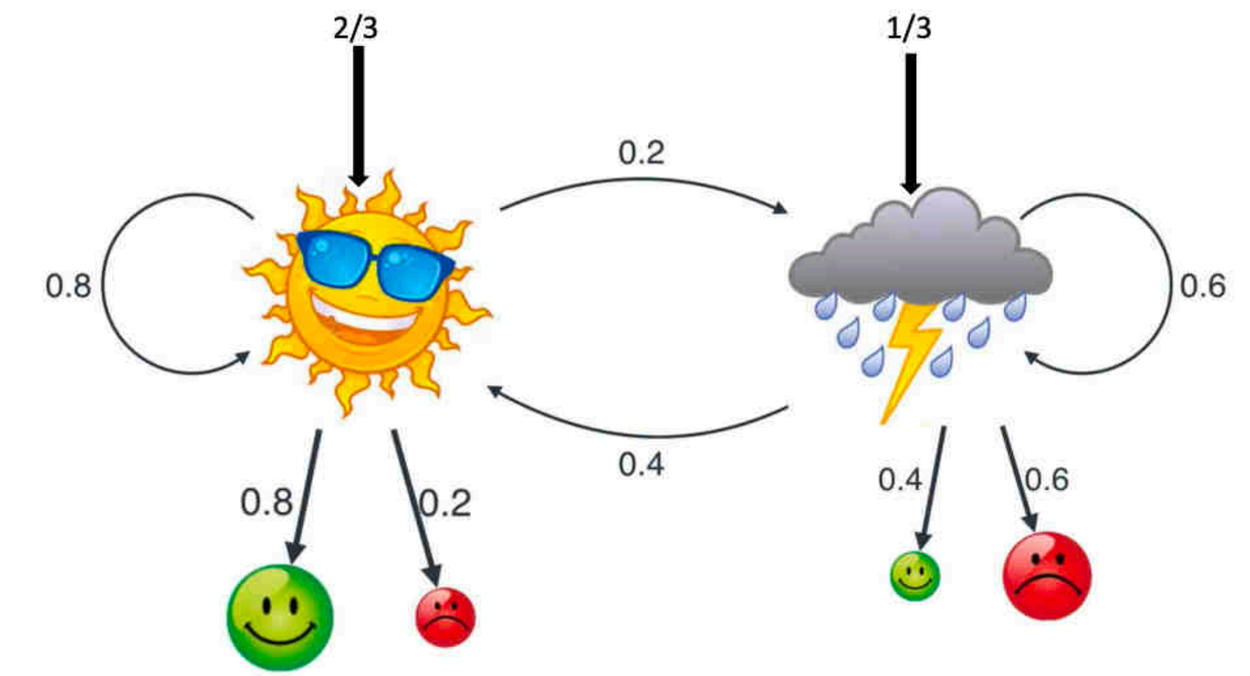
$$P(q_n = a | q_1 q_2 \dots q_{n-1}) = P(q_n = a | q_{n-1})$$

(2) Output Independence Assumption

$$P(o_i | q_1 q_2 \dots q_{i-1}) = P(o_i | q_i)$$

output is dependent only on current state

HMM example 1



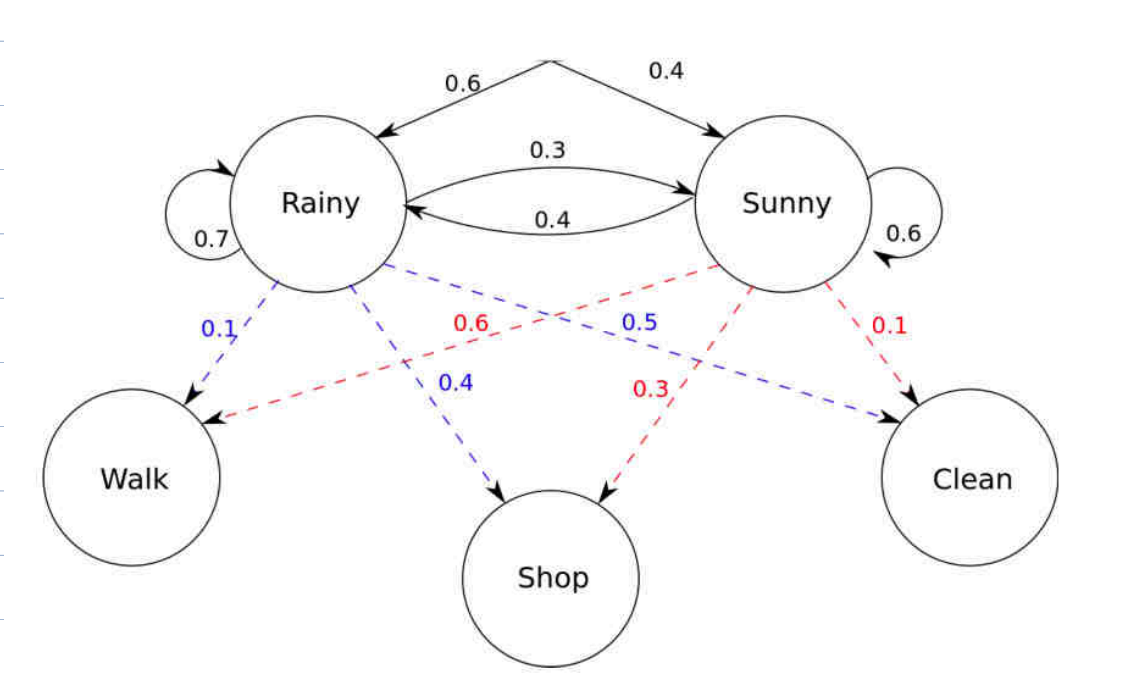
$$\pi = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$A = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

$$\lambda(\pi, A, B)$$

HMM Example 2



$$\pi = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.7 & 0 \cdot 3 \\ 0.4 & 0.6 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.6 & 0.3 & 0.1 \end{bmatrix}$$

2.1 The Likelihood Problem

Given:

Observation sequence O

HMM $\lambda(\pi, A, B)$

To find:

Likehood of observing O $P(O|\lambda)$

(a) Brute Force Solution

- Let $O = \{o_1, o_2, \ldots, o_T\}$
- Let state sequence = $\{q_1, q_2, q_3, \ldots, q_T\}$
- Probability of observing o_i from a state sequence $q_1 q_2 \ldots q_i$

$$P\left(o_i | q_1 q_2 \ldots q_i\right) = P\left(o_i | q_i\right) \times P\left(q_i | q_1 q_2 \ldots q_{i-1}\right)$$

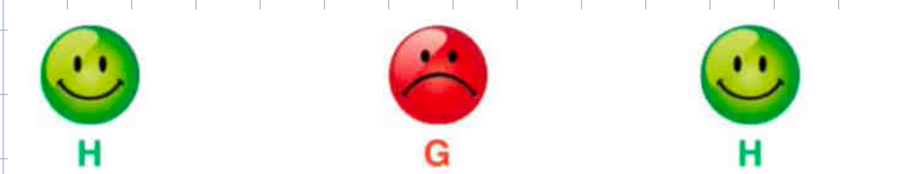
- Probability of observing O from Q

$$P(O|Q) = \prod_{i=1}^{T} P\left(o_i | q_i\right) \times P\left(q_i | q_{i-1}\right)$$

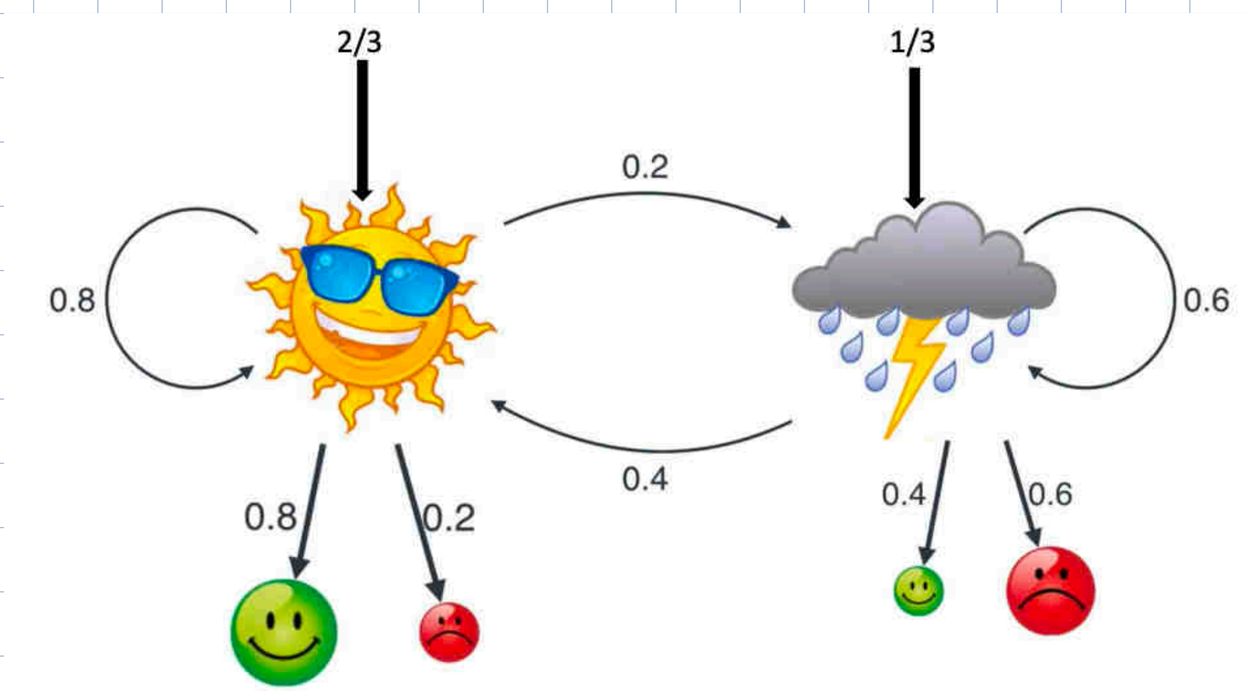
- $P(O|\lambda)$ (all possible Q sequences)

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

Q: For observation sequence O

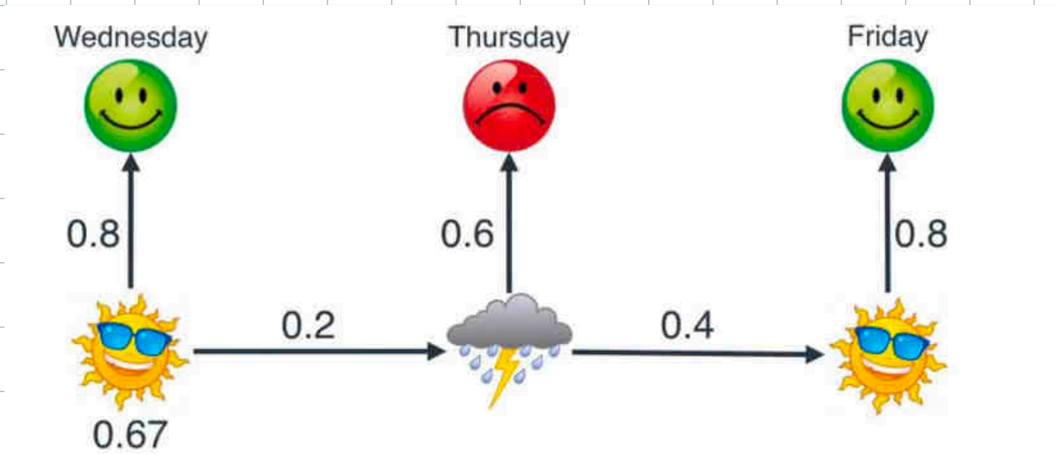


and HMM λ



Find $P(O|\lambda)$ using Brute force

All possible Q state sequences



$$P\,(HGH|SRS) = \frac{2}{3} \times 0.8 \times 0.2 \times 0.6 \times 0.4 \times 0.8$$

$$= 0.02048$$

Need to solve $2^3 = 8$ times (exponential)

Q: If an HMM has 4 hidden states, $O = \{o_1, o_2, o_3, o_4, o_5\}$, how many state sequences of length 5 can generate the given obs sequence of length 5?

Observation sequence = $o_1 o_2 o_3 o_4 o_5$

For each of the 5 emissions, 4 states could have caused it

$$4^5 = 1024$$

In general, N states, T observations

N^T possibilities

Q: Total no. of operations in brute force?

N^T sequences

$2T$ multiplications per sequence

(1 for $P(q_i | q_{i-1})$ and 1 for $P(o_i | q_i)$)

one final N^T to sum over all sequences

$$(2T+1) N^T$$

(b) Forward Algorithm

- Compute likelihood till the $t-1$ timestamp
- Update likelihood at time t based on observation o_t
- Let $\alpha_t(j)$ = joint probability of observing observations $o_1, o_2, \ldots, o_t$ and reaching state q_j at time t

$$\alpha_t(j) = P(o_1, o_2, \ldots, o_t, q_t = j|\lambda)$$

$$= \sum_{i=1}^{N} P(o_1, o_2, \ldots, o_t, q_{t-1} = i, q_t = j|\lambda)$$

Simplifying $o_1, o_2, \ldots, o_{t-1} = O_{1,t-1}$

$$= \sum_{i=1}^{N} P(O_{1,t-1}, q_{t-1} = i)\ P(q_t = j|q_{t-1} = i, O_{1,t-1})\ P(o_t|q_t = j, q_{t-1} = i, O_{1,t-1})$$

$$= \sum_{i=1}^{N} P(O_{1,t-1}, q_{t-1} = i)\ \underbrace{P(q_t = j|q_{t-1} = i)}_{\text{independent of } O_{1,t-1}}\ \underbrace{P(o_t|q_t = j, q_{t-1} = i)}_{\text{independent of } O_{1,t-1}}$$

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(o_t)$$

Initialisation

$$\alpha_1(i) = \pi_i b_i(o_1)$$

Recursion

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(o_t)$$

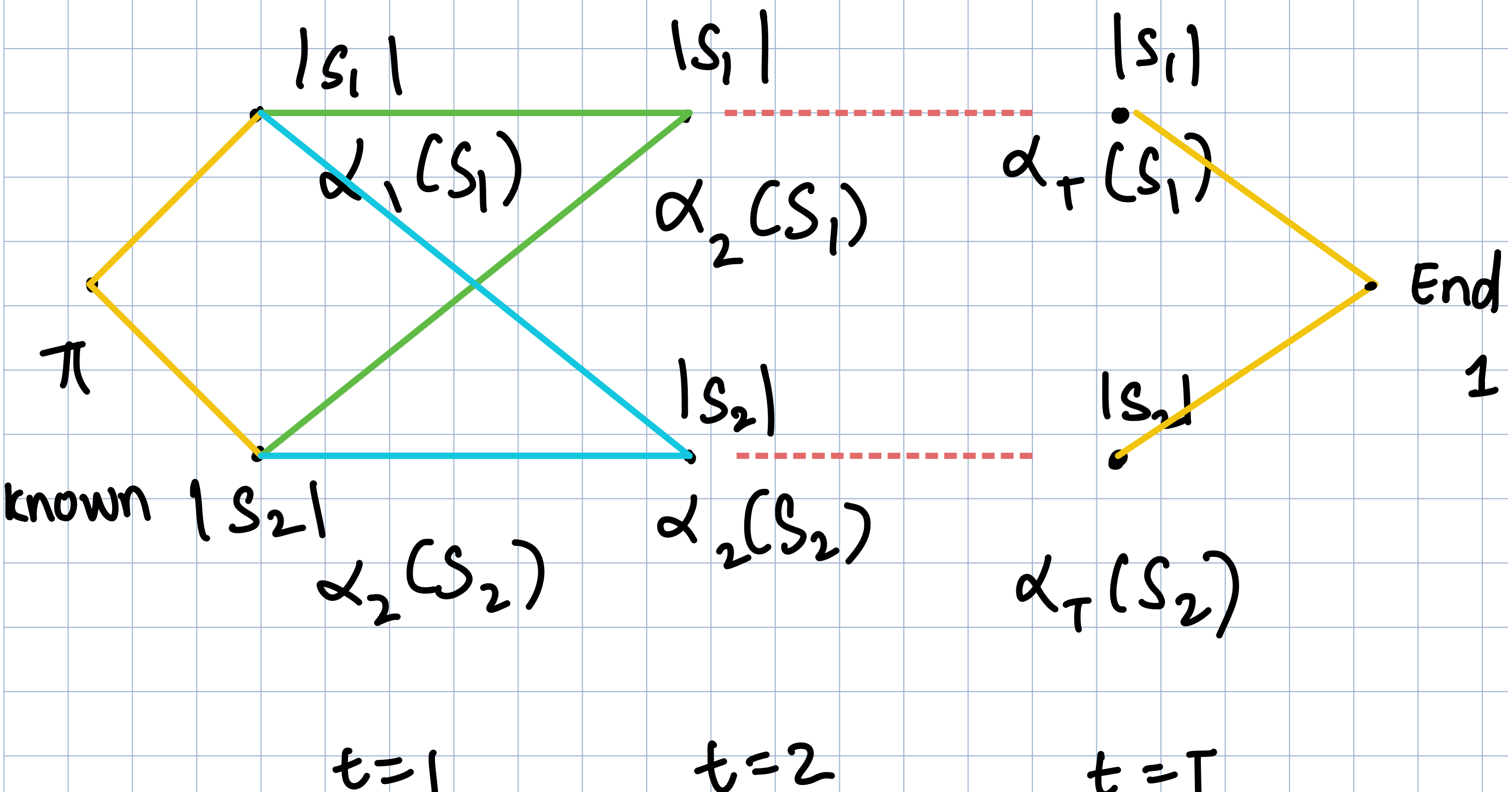
Termination

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

Algorithm

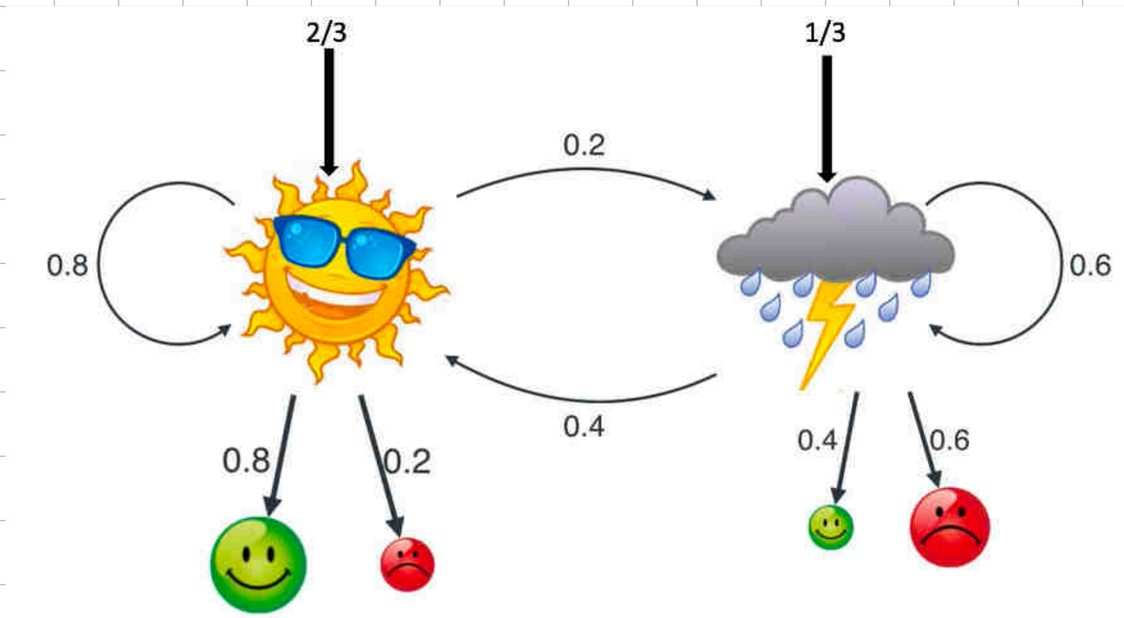
Algorithm 1 Forward Algorithm

procedure FORWARD(observations of len T, state-graph of len N)
    $forward \leftarrow$ **new** matrix[N][T]
    for state s **from** 1 to N **do**
        forward[s, 1] $\leftarrow \pi_s b_s(o_1)$
    for timestamp t from 2 to T **do**
        for state s **from** 1 to N **do**
            forward[s, t] $\leftarrow \left(\sum_{s'=1}^{N} forward[s', t] \times a_{s's}\right) \times b_s(o_t)$
    forwardProb $\leftarrow \sum_{s=1}^{N} forward[s, T]$
    return forwardProb



Q: For observation sequence O

and HMM λ



Find $P(O|\lambda)$ using Forward Algorithm

	T = 1 (Happy)	T = 2 (Grumpy)	T=3 (Happy)
Sunny	2/3 x 0.8 = 8/15	12/125	304/3125
Rainy	1/3 x 0.4 = 2/15	14/125	108/3125

$$\alpha_2(Sunny) = \left(\frac{8}{15} * 0.8 * 0.2\right) + \left(\frac{2}{15} * 0.4 * 0.2\right) = \frac{12}{125}$$

$$\alpha_3(Sunny) = \left(\frac{12}{125} * 0.8 * 0.8\right) + \left(\frac{14}{125} * 0.4 * 0.8\right) = \frac{304}{3125}$$

$$\alpha_2(Rainy) = \left(\frac{8}{15} * 0.2 * 0.6\right) + \left(\frac{2}{15} * 0.6 * 0.6\right) = \frac{14}{125}$$

$$\alpha_3(Rainy) = \left(\frac{12}{125} * 0.2 * 0.4\right) + \left(\frac{14}{125} * 0.6 * 0.4\right) = \frac{108}{3125}$$

$P(O|\lambda) = 0.13184$

(c) Backward Algorithm

- Likelihood of observing from $t+1$ to T
- Let $\beta_t(j)$ be probability that future observations $o_{t+1}, o_{t+2}, \ldots, o_T$ have been observed given that hidden state at t is j

$$\beta_t(j) = P(o_{t+1}, o_{t+2}, \ldots, o_T | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(o_{t+1}, o_{t+2}, \ldots, o_T, q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(o_{t+1}, O_{t+2,T}, q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(o_{t+1}, O_{t+2,T} | q_{t+1} = i, q_t = j, \lambda)\ P(q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(o_{t+1}, O_{t+2,T} | q_{t+1} = i, q_t = j, \lambda)\ a_{ji}$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+2,T} | o_{t+1}, q_{t+1} = i, q_t = j, \lambda)\ P(o_{t+1} | q_{t+1} = i, q_t = j)\ a_{ji}$$

$$\beta_t(j) = \sum_{i=1}^{N} \beta_{t+1}(i)\ b_i(o_{t+1})\ a_{ji}$$

Initialisation

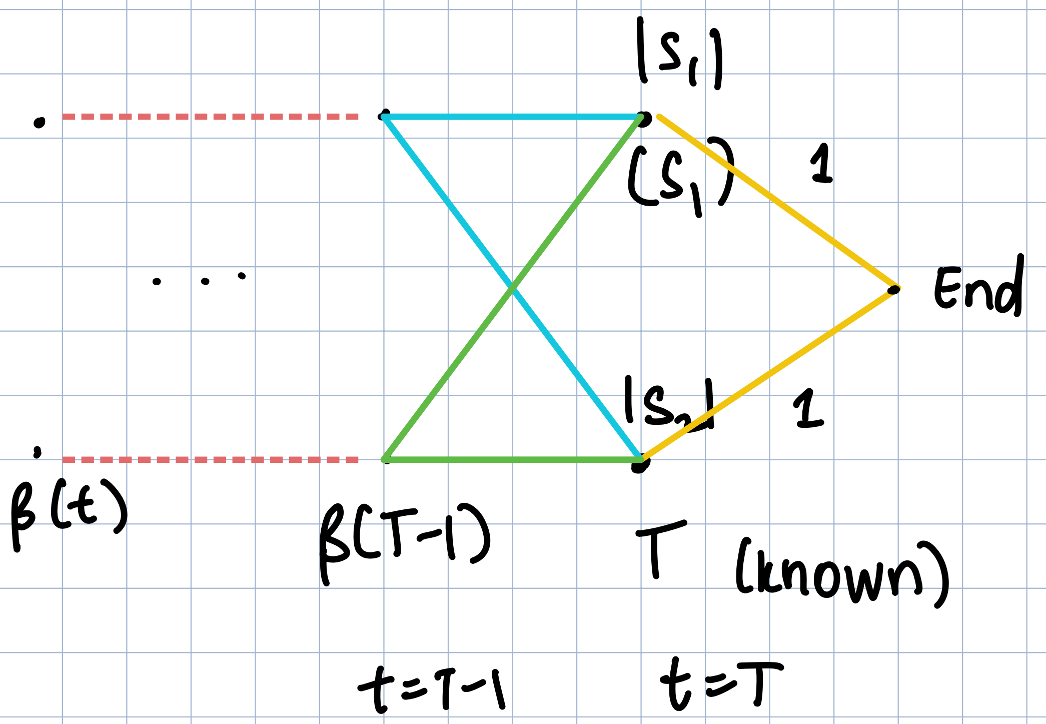
$$\beta_T(i) = 1$$

Recursion

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

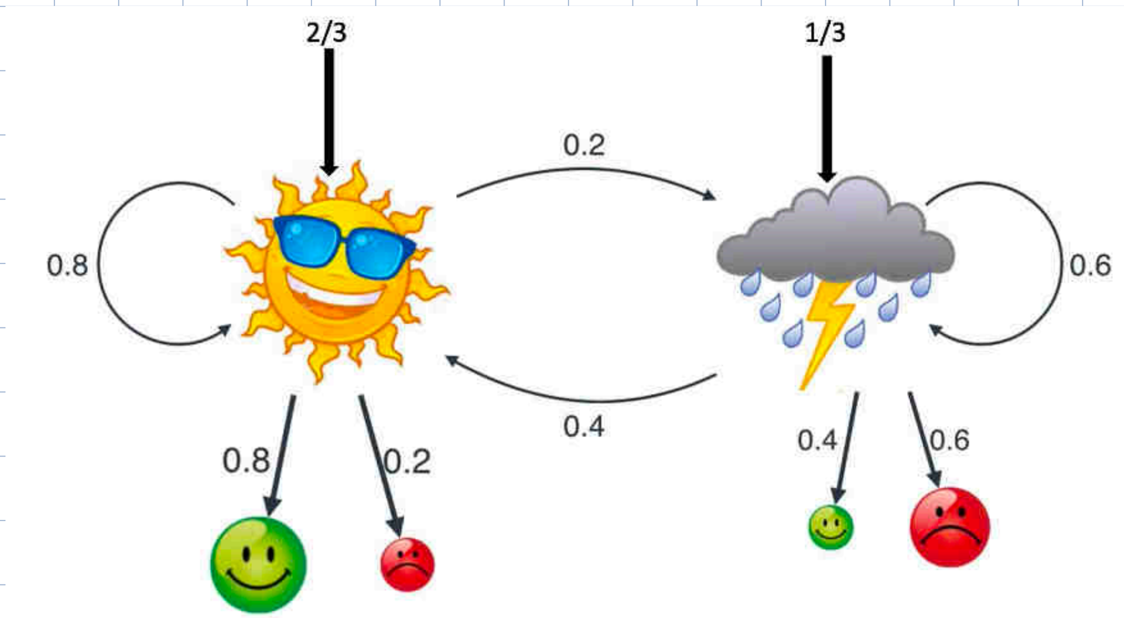
Termination

$$P(O|\lambda) = \sum_{j=1}^{N} \pi_j b_j(O_1) \beta_1(j)$$



Q: For observation sequence O

and HMM λ



Find P(O|λ) using Backward Algorithm

	T = 1 (Happy)	T = 2 (Grumpy)	T=3 (Happy)
Sunny	0.1824	0.72	1
Rainy	0.2592	0.56	1

$$\beta_2(Sunny) = (0.8 * 0.8 * 1) + (0.2 * 0.4 * 1) = 0.72$$

$$\beta_1(Sunny) = (0.8 * 0.2 * 0.72) + (0.2 * 0.6 * 0.56) = 0.1824$$

$$\beta_2(Rainy) = (0.4 * 0.8 * 1) + (0.6 * 0.4 * 1) = 0.56$$

$$\beta_1(Rainy) = (0.4 * 0.2 * 0.72) + (0.6 * 0.6 * 0.56) = 0.2592$$

P(O|λ) = 0.1318

Important Result

$$\alpha_t(j) = P(o_1, o_2, \ldots, o_t, q_t = j|\lambda)$$

$$\beta_t(j) = P(o_{t+1}, o_{t+2}, \ldots, o_T|q_t = j, \lambda)$$

$\therefore$ $\alpha_t(j) \times \beta_t(j)$ = prob of being in j at t given O

$\therefore$ dot product of i^{th} columns of forward and backward matrix should be same for all columns

2.2 Decoding Problem

Given: sequence of observations O
HMM $\lambda(\pi, A, B)$

To find: most likely sequence of states Q
that emitted O

$$Q_{best} = \arg\max_{Q} P\left(O|Q\right)$$

(a) Brute Force

- Similar to Likelihood Problem

(b) Viterbi Algorithm

- Define

$$\nu_t\left(j\right) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_{t-1}, o_1, o_2, \ldots, o_{t-1}, q_t = j, o_t|\lambda)$$

- Back pointers bt_t used

Initialisation

$$\nu_1(j) = \pi_j b_j(o_1)$$

$$bt_1(j) = 0$$

Recursion

$$\nu_t(j) = \max_{i=1}^{N} \nu_{t-1}(i)\, a_{ij} b_j(o_t)$$

$$bt_t(j) = \arg\max_{i=1}^{N} \nu_{t-1}(i)\, a_{ij} b_j(o_t)$$

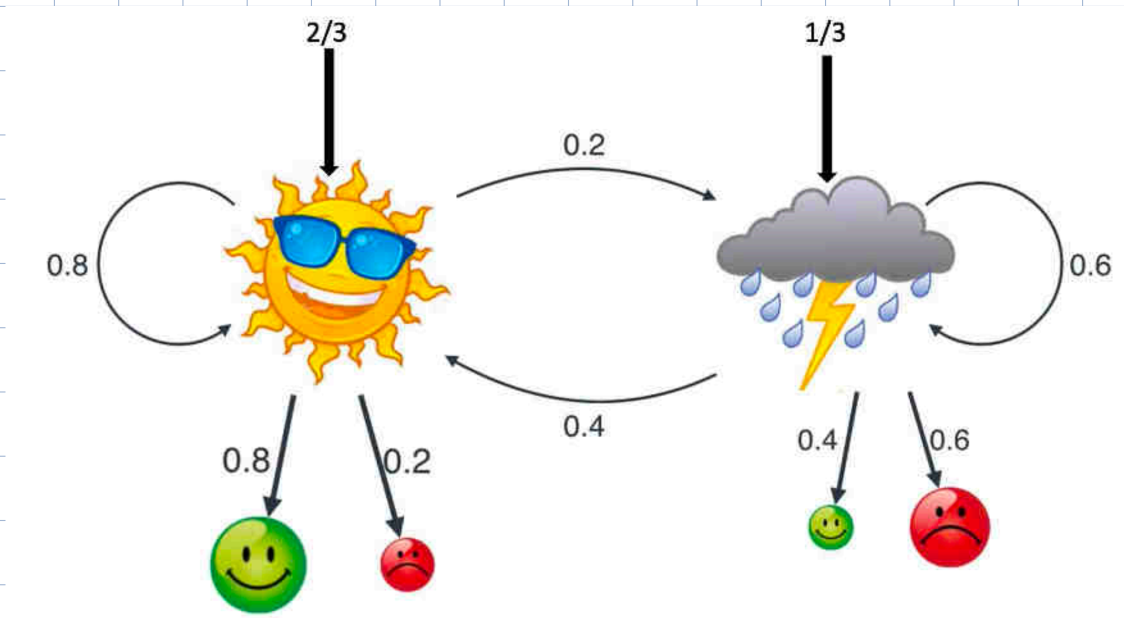
Termination

$$\text{Best score}: P* = \max_{i=1}^{N} \nu_T(i)$$

$$\text{Best path start}: q_T* = \arg\max_{i=1}^{N} \nu_T(i)$$

Q: For observation sequence O

and HMM λ



Find most likely seq using Viterbi Algorithm

	T = 1 (Happy)	T = 2 (Grumpy)	T=3 (Happy)
Sunny	2/3 x 0.8 = 8/15	32/375	512/9375
Rainy	1/3 x 0.4 = 2/15	8/125	144/9375

$$v_2(Sunny) = \max\left(\left(\frac{8}{15} * 0.8 * 0.2\right), \left(\frac{2}{15} * 0.4 * 0.2\right)\right) = \frac{32}{375}$$

$$v_3(Sunny) = \max\left(\left(\frac{32}{375} * 0.8 * 0.8\right), \left(\frac{8}{125} * 0.4 * 0.8\right)\right) = \frac{512}{9375}$$

$$v_2(Rainy) = \max\left(\left(\frac{8}{15} * 0.2 * 0.6\right), \left(\frac{2}{15} * 0.6 * 0.6\right)\right) = \frac{8}{125}$$

$$v_3(Rainy) = \max\left(\left(\frac{32}{375} * 0.2 * 0.4\right), \left(\frac{8}{125} * 0.6 * 0.4\right)\right) = \frac{144}{9375}$$

2.3 Learning Problem

Given: Observation sequence O
set of hidden states Q

To find: A and B

(a) Baum-Welch Algorithm

- Special case of EM (iteratively estimates)

- Estimating a_{ij} and $b_j(k)$

$$\hat{a}_{ij} = \frac{\text{expected no. of trans from } i \text{ to } j}{\text{expected no. of trans from } i}$$

$$\hat{b}_j(k) = \frac{\text{expected no. of times in } j \text{ observing } k}{\text{expected no. of times in } j}$$

We know

$$P(X|Y,Z) = \frac{P(X,Y|Z)}{P(Y|Z)}$$

For $\hat{a}_{ij}$

- Define $\xi_t(i,j)$ (xi) as probability of being in state i at time t and state j at time $t+1$ given the observation seq O and the model λ

$$\xi_t(i,j) = P(q_t = i, q_{t+1} = j | O, \lambda)$$

- Define

$$\text{almost-}\xi_t(i,j) = P(q_t = i, q_{t+1} = j, O | \lambda)$$

(includes probability of observing O)

$$P(q_t = i, q_{t+1} = j, O|\lambda) = \alpha_t(i)\, a_{ij} b_j(o_{t+1})\, \beta_{t+1}(j)$$

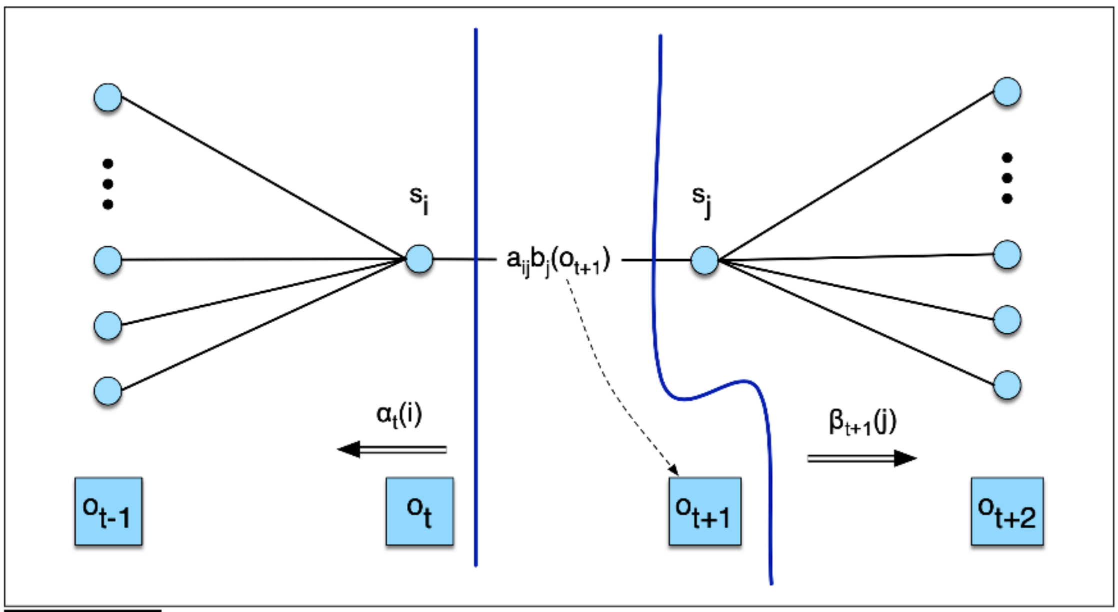


Figure A.12 Computation of the joint probability of being in state i at time t and state j at time $t+1$. The figure shows the various probabilities that need to be combined to produce $P(q_t = i, q_{t+1} = j, O|\lambda)$: the α and β probabilities, the transition probability a_{ij} and the observation probability $b_j(o_{t+1})$. After Rabiner (1989) which is ©1989 IEEE.

$$P(q_t = i, q_{t+1} = j | O, \lambda) = \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)}$$

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_t(j)\beta_t(j)$$

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

Total expected no. of transitions from state i to state j is summation of $\xi_t(i,j)$ from 1 to $T-1$

Total expected no. of transitions starting from state i : summing over all values of state k from 1 to N

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)} \rightarrow (1)$$

For $\hat{b}_j(k)$

- Define

$$\gamma_t(j) = P(q_t = j | O, \lambda)$$

$$\gamma_t(j) = \frac{P(q_t = j, O | \lambda)}{P(O|\lambda)}$$

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_t(j)\beta_t(j)$$

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$$

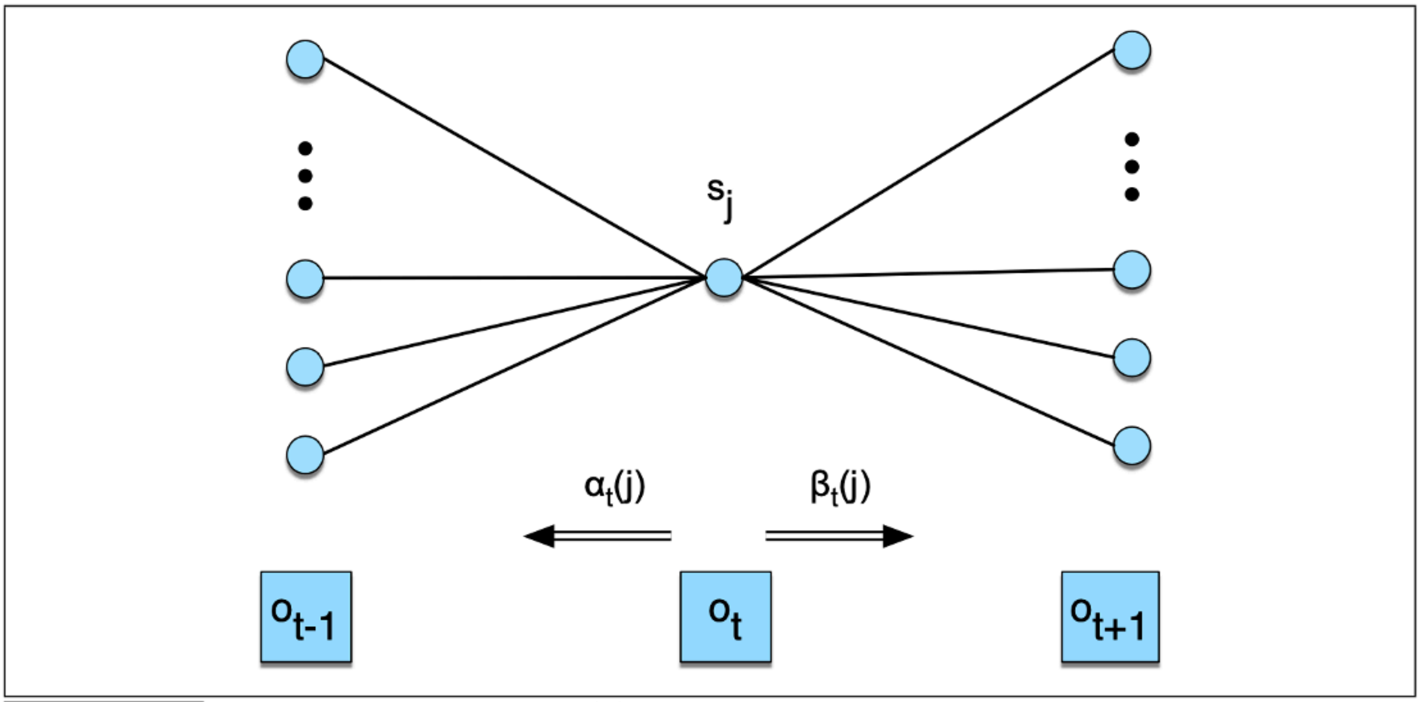


Figure A.13 The computation of $\gamma_t(j)$, the probability of being in state j at time t. Note that γ is really a degenerate case of ξ and hence this figure is like a version of Fig. A.12 with state i collapsed with state j. After Rabiner (1989) which is ©1989 IEEE.

$$\hat{b}_j(k) = \frac{\sum_{t=1\ s.t.\ o_t=k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \quad \text{—(2)}$$

Equations (1) and (2) – re-estimate A and B

Start with initial estimates for A & B

(1) E step: compute expected γ and ξ from initial A & B

(2) M step: recompute A & B from γ and ξ

function FORWARD-BACKWARD(*observations* of len *T*, *output vocabulary V*, *hidden state set Q*) **returns** *HMM=(A,B)*

initialize *A* and *B*
iterate until convergence
E-step

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\alpha_T(q_F)} \ \forall\, t \text{ and } j$$

$$\xi_t(i,j) = \frac{\alpha_t(i)\,a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\alpha_T(q_F)} \ \forall\, t,\ i, \text{ and } j$$

M-step

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)}$$

$$\hat{b}_j(v_k) = \frac{\sum_{t=1 s.t.\ O_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

return *A*, *B*

If K sequences given, k^{th} sequence is of length T_k

Assuming each sequence is independent

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k-1} \xi_t^k(i,j)}{\sum_{k=1}^{K} \sum_{t=1}^{T_k-1} \gamma_t^k(i)}$$

$$\hat{b}_j(m) = \frac{\sum_{k=1}^{K} \sum_{t=1 \text{ s.t. } o_t=m}^{T_k} \gamma_t(j)}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_t(j)}$$

$$\hat{b}_j(m) = \frac{\sum_{k=1}^{K} \sum_{t=1 \text{ s.t. } o_t=m}^{T_k} \gamma_t(j)}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_t(j)}$$

$$\hat{\pi}_i = \frac{\sum_{k=1}^{K} \gamma_1^k(i)}{K}$$

Pset

One of the most famous examples in HMM is that of the *dishonest casino owner*. The internal state of the HMM denotes whether a casino owner is using a fair die or a loaded die (i.e. an unfair one). The observations denote the number shown on the die (1 to 6).

It is equally likely that the first roll is made from the fair or loaded die. The transition and emission probabilities are:

	Fair	**Loaded**
Fair	0.99	0.01
Loaded	0.2	0.8

Transition Matrix for Dishonest Casino

	1	**2**	**3**	**4**	**5**	**6**
Fair	1/6	1/6	1/6	1/6	1/6	1/6
Loaded	1/10	1/10	1/10	1/10	1/10	1/2

Emission Matrix for Dishonest Casino

Find the order of dice that was most likely used by the owner for 3 rolls, given that the numbers that came up on those 3 rolls of the dice are $6, 2, 6$.

	6	2	6
Fair	$1/2 \times 1/6 = 1/12$	$11/800$	2.27×10^{-3}
Loaded	$1/2 \times 1/2 = 1/4$	$1/50$	8×10^{-3}

$\max(1/12 \times 0.99 \times 1/6, \; 1/4 \times 0.2 \times 1/6)$

$\max(11/800, \; 1/120)$

$\max(0.01375, \; 8 \times 10^{-3})$

$\max(1/12 \times 0.01 \times 1/10, \; 1/4 \times 0.8 \times 1/10)$

$\max(1/12000, \; 1/50)$

$\max(11/800 \times 0.99 \times 1/6, \frac{1}{50} \times 0.2 \times 1/6)$

$\max(2.27 \times 10^{-3}, 6.67 \times 10^{-4})$

$\max(\frac{11}{800} \times 0.01 \times \frac{1}{2}, \frac{1}{50} \times 0.8 \times \frac{1}{2})$

$\max(6.8 \times 10^{-5}, 8 \times 10^{-3})$

sequence: Load → Load → Load

. One biological application of HMMs is to determine the secondary structure (i.e. the general 3D shape) of a protein. This general shape is made up of alpha helices, beta sheets, and other structures. Assume that the amino acid composition of these regions (in terms of 6 amino acids M, L, N ,E, A and G) is governed by an HMM.

The start state is always "other". The emission and transition probabilities are:

	Alpha	Beta	Other
Alpha	0.7	0.1	0.2
Beta	0.2	0.6	0.2
Other	0.3	0.3	0.4

Transition Matrix for Protein Structure

	M	L	N	E	A	G
Alpha	0.35	0.30	0.15	0.10	0.05	0.05
Beta	0.10	0.05	0.30	0.40	0.00	0.15
Other	0.05	0.15	0.20	0.15	0.20	0.25

Emission Matrix for Protein Structure

i. How many paths could give rise to the sequence $O = MLN$? What is the total probability $P(O)$?

ii. Give the most likely state transition path $q*$ for the amino acid sequence MLN

$\pi = (0, 0, 1)$

(i) No of paths = $1 \times 3 \times 3 = 9$

Total probability

	M	L	N
Other	$1 \times 0.05 = 0.05$	3×10^{-3}	4.5×10^{-4}
Alpha	0	4.5×10^{-3}	6.3×10^{-4}
Beta	0	7.5×10^{-4}	5.4×10^{-4}

$P(O) = 1.62 \times 10^{-3}$

(ii)

	M	L	N
Other	$1 \times 0.05 = 0.05$	3×10^{-3}	2.4×10^{-4}
Alpha	0	4.5×10^{-3}	4.725×10^{-4}
Beta	0	7.5×10^{-4}	1.35×10^{-4}

$$\max(1.2\times10^{-3}, 9\times10^{-4}, 1.5\times10^{-4}) \times 0.2$$

$$\max(9\times10^{-4}, 3.15\times10^{-3}, 1.5\times10^{-4}) \times 0.15$$

$$\max(9\times10^{-4}, 4.5\times10^{-4}, 4.5\times10^{-4}) \times 0.3$$

sequence : Other - Alpha - Alpha