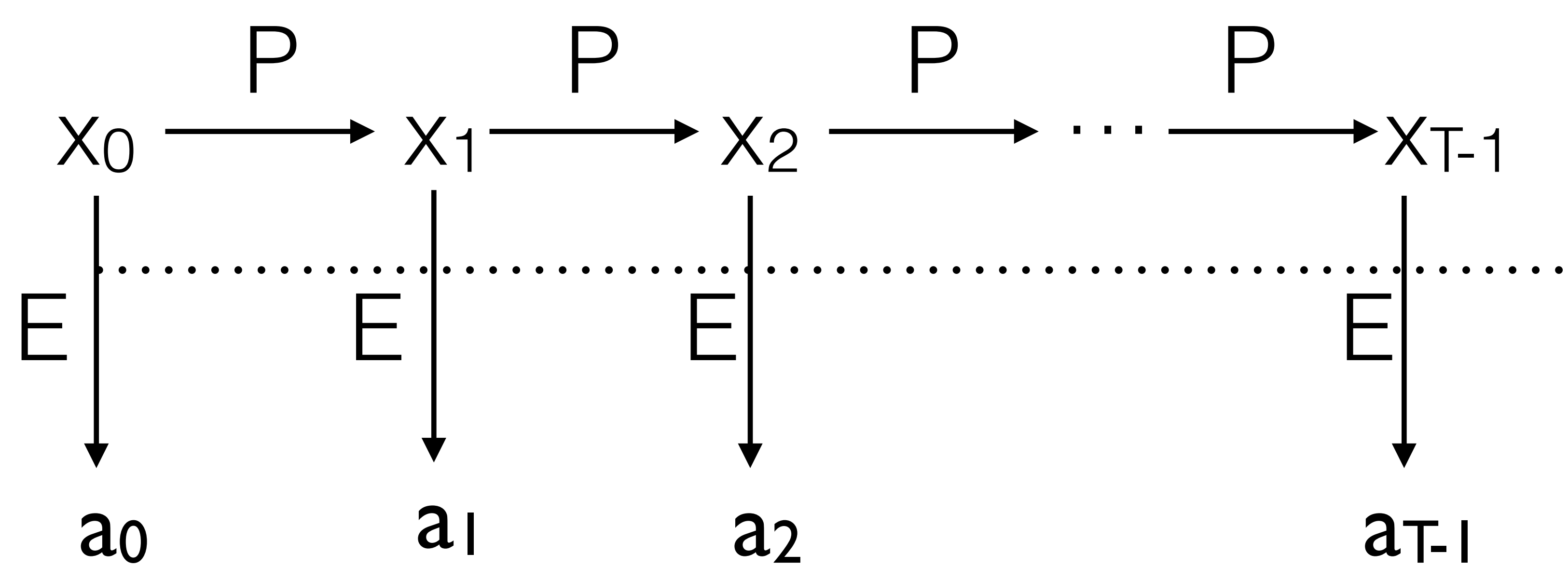
Hidden Markov Models

Juliana Silva Bernardes

Hidden Markov Models

- Formal definition:
  - A hidden Markov model is a Markov process $\lambda = (Q, P, E, \pi)$

  where Q is the set of distinct states, P state transition probabilities, E is the observation probability matrix and π is the initial state distribution.



O = observed sequence

T = length of the observed sequence $|O| = T$

N = number of states in the model, $|Q| = N$

V = set of observable symbols

M = number of observable symbols $|V| = M$

Hidden Markov Models

- Example : average annual temperature*
  - Suppose we want to determine the average annual temperature at a particular location on earth over a series of years
  - To make it interesting, suppose the years we are concerned with lie in the distant past, before thermometers were invented.
  - Since we can't go back in time, we instead look for indirect evidence of the temperature.
  - For simplicity, we only consider two annual temperatures, "hot" and "cold"
  - Suppose that modern evidence indicates that the probability are:

$$P= \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \left[\begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array}\right] \end{array}$$

*http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

Hidden Markov Models

- Example : average annual temperature*

  - Also suppose that current research indicates a correlation between the size of tree growth rings and temperature
  - For simplicity, we only consider three different tree ring sizes, small, medium and large, or S, M and L,
  - The probabilistic relationship between annual temperature and tree ring sizes is given by

$$E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

*http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

Hidden Markov Models

- Example : average annual temperature*

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \left[\begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array}\right] \end{array} \qquad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \left[\begin{array}{ccc} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{array}\right] \end{array}$$

- For this system, the state is the average annual temperature (H or C).
- The transition from one state to the next is a Markov process (of order one), since the next state depends only on the current state and the fixed probabilities P.
- However, the actual states are "hidden" since we can't directly observe the temperature in the past.
- we can't observe the state (temperature) in the past, but we can observe the size of tree rings at regarding E.
- Since the states are hidden, this type of system is known as a Hidden Markov Model (HMM).
- Our goal is to make effective and efficient use of the observable information so as to gain insight into various aspects of the Markov process.

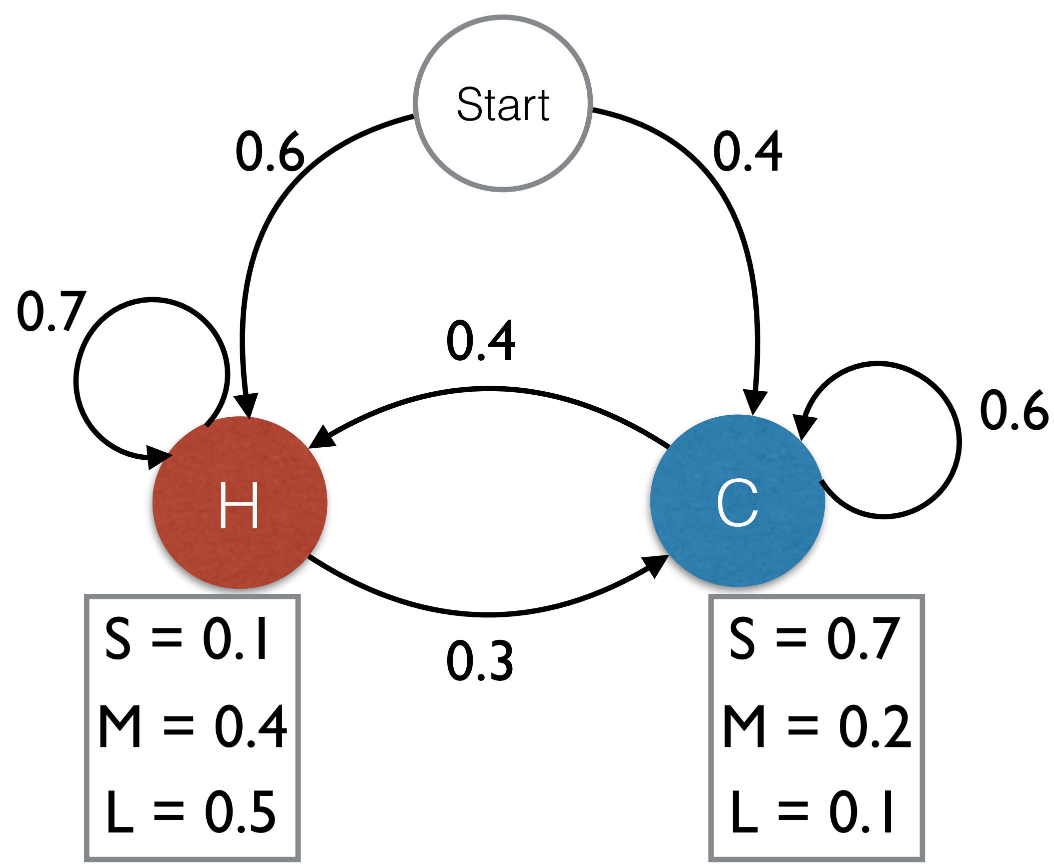
*http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

Hidden Markov Models

- Graph representation of λ =(Q, P, E, π)

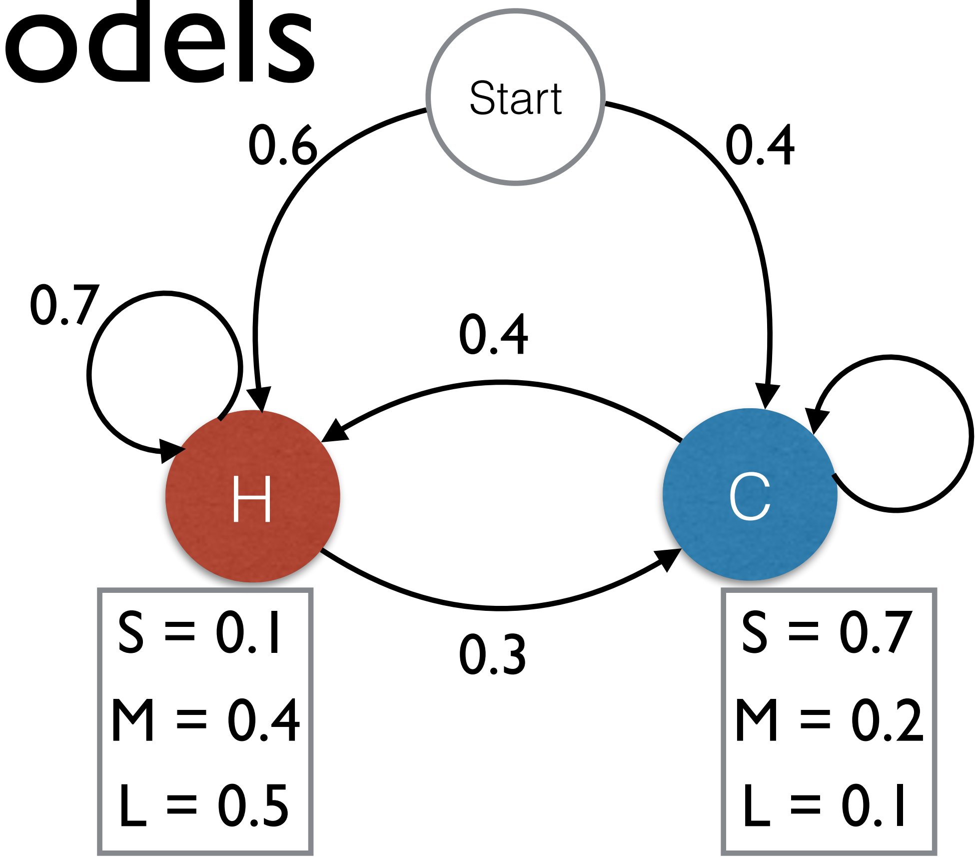
$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

- suppose that the initial state distribution, denoted by π_0, is π $\begin{array}{cc} H & C \\ [0.6 & 0.4] \end{array}$



*http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

Hidden Markov Models



- Now consider a particular four-year period of interest from the distant past, for which we observe the series of tree rings S,M,S,L.

O(S,M,S,L)

- We might want to determine the most likely state sequence of the Markov process given the observations

- That is, we might want to know the most likely average annual temperatures over the four-year period of interest

*http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

Hidden Markov Models

$\lambda = (Q, P, E, \pi)$

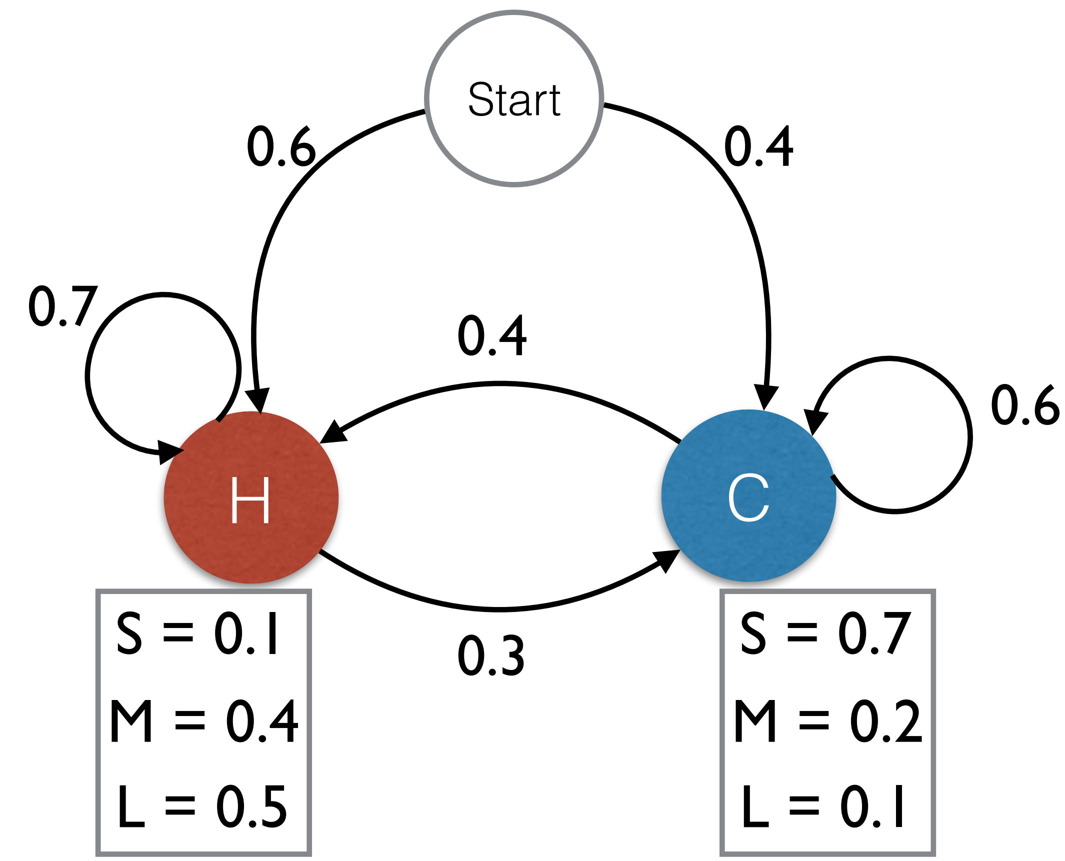
Q is the set of distinct states, $Q = \{H, C\}$

N = number of states in the model, $N = 2$

V = set of observable symbols $V = \{S, M, L\}$

M = number of observable symbols $M = 3$

Given the observation $O(S,M,S,L)$ $T = 4$



Hidden Markov Models

$\lambda = (Q, P, E, \pi)$

- The matrix $P = (p_{ij})$ is N x N with

  $p_{ij} = P$ (state x_j at t+1 | state x_i at t)

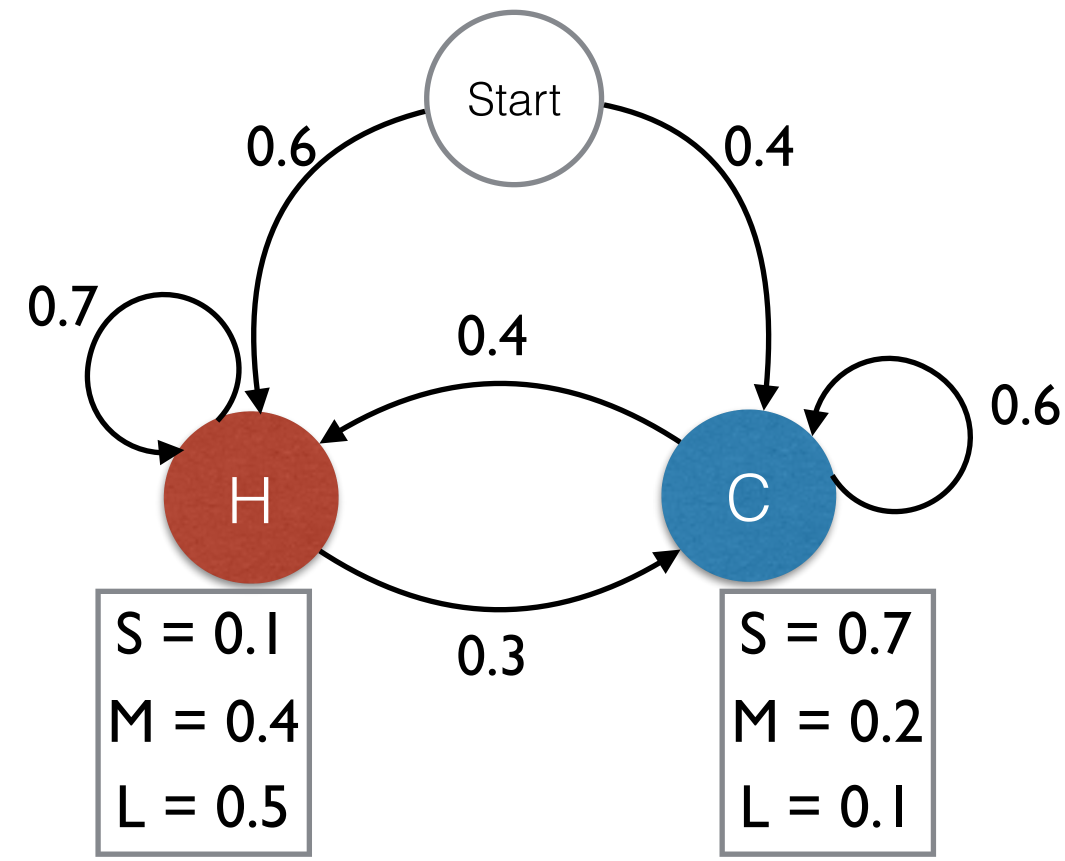
- The matrix $E = (e_i)$ is N x M with

  $e_i(a) = E$ (observation a at t | state x_i at t)

For simplicity we can say a HMM is defined by: $\lambda = (P, E, \pi)$

Hidden Markov Models

- Now, consider a generic state sequence of length four $X = \{x_0, x_1, x_2, x_3\}$
  - With corresponding observations $O = \{a_0, a_1, a_2, a_3\}$



π_{x0} is the probability of starting in state x_0

$e_{x0}(a_0)$ is the probability of initially observing a_0

$p_{x0,x1}$ is the probability of transiting from state x_0 to state x_1

- Then the probability of the state sequence X and the observation O according to λ is given by

$$P(X, O) = \pi_{x_0} e_{x_0}(a_0)\ p_{x_0,x_1} e_{x_1}(a_1)\ p_{x_1,x_2} e_{x_2}(a_2)\ p_{x_2,x_3} e_{x_3}(a_3)$$

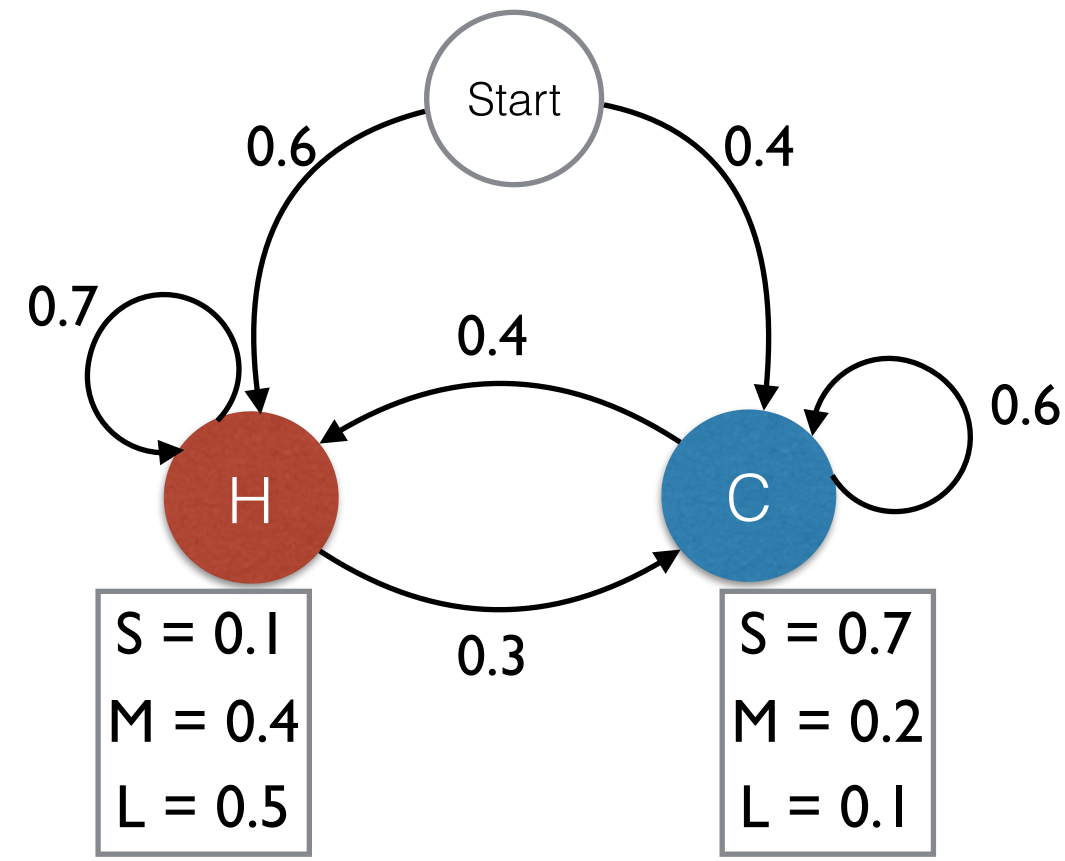
$$P(HHCC, SMSL) = 0.6(0.1)\ 0.7(0.4)\ 0.3(0.7)\ 0.6(0.1) = 0.000212$$

$$\boxed{P(X,O) = \pi_{x_0} e_{x_0}(a_0)\ \Pi\ p_{x_{t-1},x_t}\ e_{x_t}(a_t)}$$

Hidden Markov Models

- Similarly, we can directly compute the probability of each possible state sequence of length four, assuming the given observation sequence O(S,M,S,L)

state	probability	normalized probability
$HHHH$	.000412	.042787
$HHHC$	.000035	.003635
$HHCH$	.000706	.073320
$HHCC$	.000212	.022017
$HCHH$	.000050	.005193
$HCHC$	.000004	.000415
$HCCH$	.000302	.031364
$HCCC$	.000091	.009451
$CHHH$	.001098	.114031
$CHHC$	.000094	.009762
$CHCH$	.001882	.195451
$CHCC$	.000564	.058573
$CCHH$	.000470	.048811
$CCHC$	.000040	.004154
$CCCH$	.002822	.293073
$CCCC$	.000847	.087963



Hidden Markov Models

- There are three fundamental problems that we can solve using HMMs
  - Problem 1:

    Given the model $\lambda = (P, E, \pi)$ and a sequence of observations O, find $P(O \mid \lambda)$.

    Here, we want to determine the likelihood of the observed sequence O, given the model.

  - Problem 2:

    Given $\lambda = (P, E, \pi)$ and an observed sequence O, find an optimal state sequence for the underlying Markov process.

  - Problem 3:

    Given an observed sequence O and the dimensions N and M, find the model $\lambda = (P,E,\pi)$ that maximizes the probability of O.

    This can be viewed as training a model to best fit the observed data

Hidden Markov Models

- Problem 1:

  Given the model λ = (P, E, π) and a sequence of observations O, find P (O | λ).

  Here, we want to determine the likelihood of the observed sequence O, given the model.

$$P(\mathcal{O}\,|\,\lambda) \;=\; \sum_X P(\mathcal{O}, X\,|\,\lambda)$$

Applying Bayes theorem $P(\mathcal{O}, X\,|\,\lambda) = \dfrac{P(\mathcal{O}\cap X\cap\lambda)}{P(\lambda)}$

$$\frac{P(\mathcal{O}\cap X\cap\lambda)}{P(\lambda)} = \underbrace{\frac{P(\mathcal{O}\cap X\cap\lambda)}{P(X\cap\lambda)}}_{P(\mathcal{O}\,|\,X,\lambda)} \cdot \underbrace{\frac{P(X\cap\lambda)}{P(\lambda)}}_{P(X\,|\,\lambda)} = P(\mathcal{O}\,|\,X,\lambda)P(X\,|\,\lambda)$$

$$\begin{aligned} P(\mathcal{O}\,|\,\lambda) &= \sum_X P(\mathcal{O}, X\,|\,\lambda) \\ &= \sum_X P(\mathcal{O}\,|\,X,\lambda)P(X\,|\,\lambda) \end{aligned}$$

Hidden Markov Models

- Problem 1:

  Given the model λ = (P, E, π) and a sequence of observations O, find P (O | λ).

  Here, we want to determine the likelihood of the observed sequence O, given the model.

$$\begin{aligned} P(\mathcal{O} \mid \lambda) &= \sum_X P(\mathcal{O}, X \mid \lambda) \\ &= \sum_X P(\mathcal{O} \mid X, \lambda) P(X \mid \lambda) \end{aligned}$$

$$P(X \mid \lambda) = P(X) = \pi_{x0}\, p_{x0,x1} \ldots p_{xT-1,xT}$$

$$P(\mathcal{O} \mid X, \lambda) = \frac{P(O, X, \lambda)}{P(X, \lambda)} = \frac{P(O, X)}{P(X)}$$

$$\boxed{P(X,O) = \pi_{x_0} e_{x_0}(a_0) \prod p_{x_{t-1},x_t} e_{x_t}(a_t)}$$

$$P(\mathcal{O} \mid X, \lambda) = \frac{P(O, X)}{P(X)} = \frac{\pi_{x0} e_{x0}(a_0)\, p_{x0,x1} e_{x1}(a_1) \ldots p_{xT-1,xT} e_T(a_T)}{\pi_{x0}\, p_{x0,x1} \ldots p_{xT-1,xT}} = e_{x0}(a_0)\, e_{x1}(a_1) \ldots e_T(a_T)$$

$$\begin{aligned} P(\mathcal{O} \mid \lambda) &= \sum_X P(\mathcal{O}, X \mid \lambda) \\ &= \sum_X P(\mathcal{O} \mid X, \lambda) P(X \mid \lambda) = \sum e_{x0}(a_0)\, e_{x1}(a_1) \ldots e_T(a_T) \quad \pi_{x0}\, p_{x0,x1} \ldots p_{xT-1,xT} \\ &= \sum \pi_{x0} e_{x0}(a_0)\, p_{x0,x1} e_{x1}(a_1) \ldots p_{xT-1,xT} e_T(a_T) \end{aligned}$$

Hidden Markov Models

- Problem 1:

  Given the model $\lambda = (P, E, \pi)$ and a sequence of observations O, find $P(O \mid \lambda)$.

  Here, we want to determine the likelihood of the observed sequence O, given the model.

$$P(O \mid \lambda) = \sum_{x} \pi_{x0} e_{x0}(a_0)\; p_{x0,x1} e_{x1}(a_1) \; \ldots \; p_{xT-1,xT} e_T(a_T)$$

- However, this direct computation is generally infeasible, since it requires about $2TN^T$ multiplications.

To find $P(O|\lambda)$ in a feasible time we use the forward algorithm

Hidden Markov Models

- Forward algorithm

Initialization: $\alpha_0(x_i) = 1$

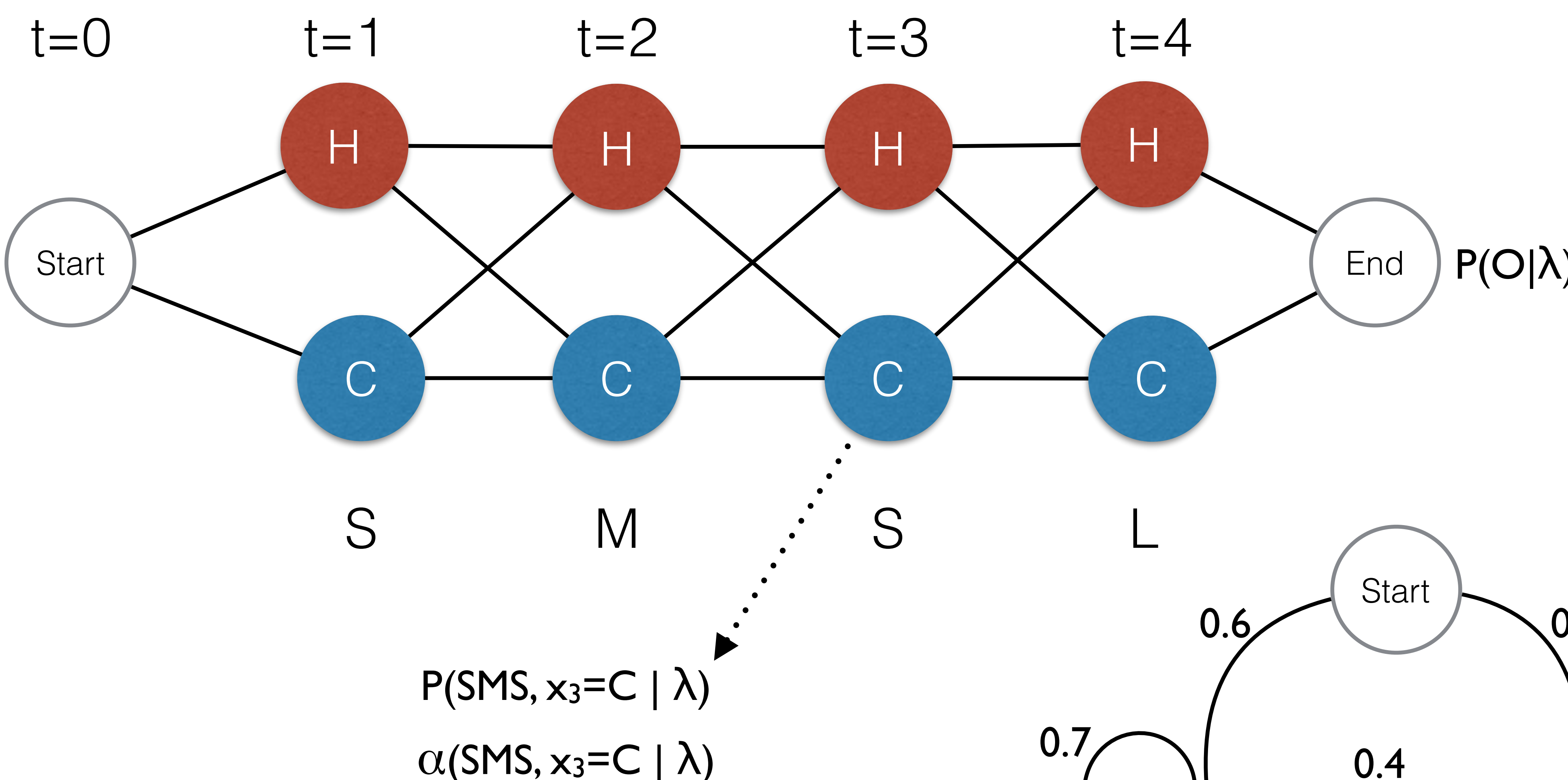
for i = 1 to T do

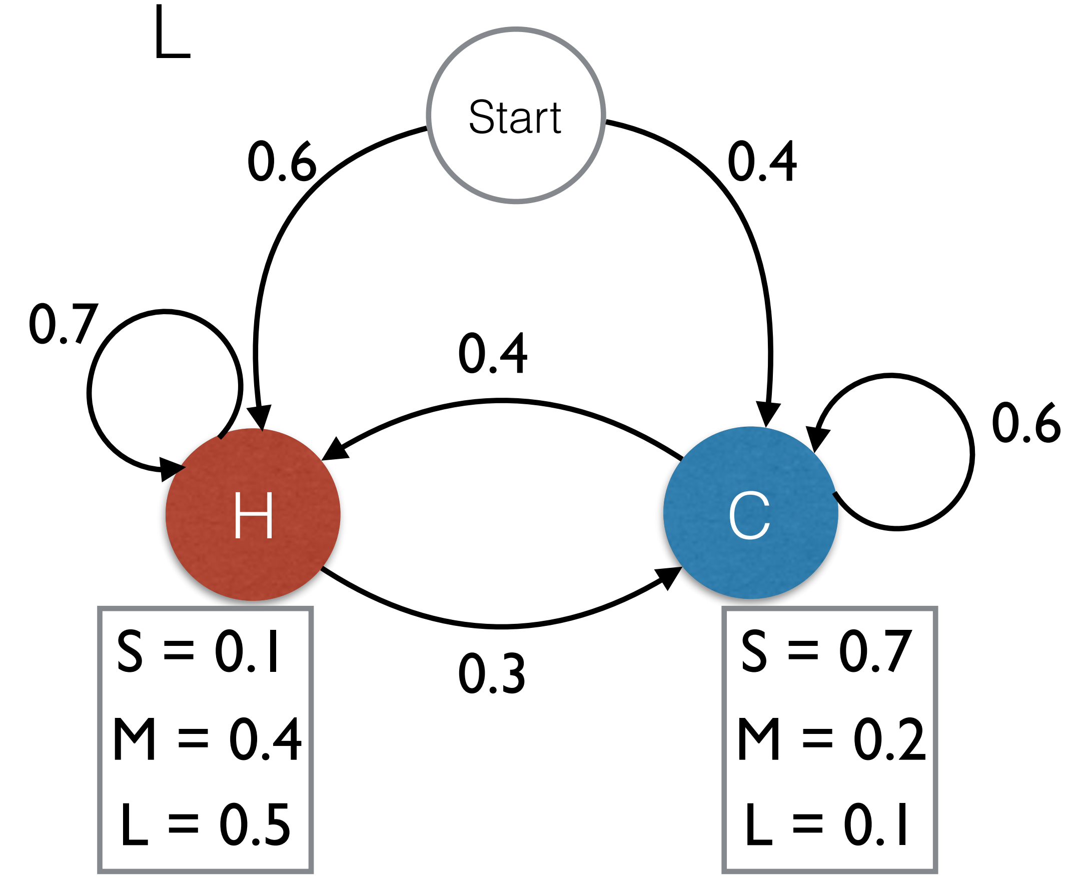
$$\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}$$

return $P(O|\lambda) = \sum_{j=1}^{N} \alpha_j(x_T)$

Hidden Markov Models

- Problem 1: To find $P(O|\lambda)$ in a feasible time we use the forward algorithm $O(S,M,S,L)$



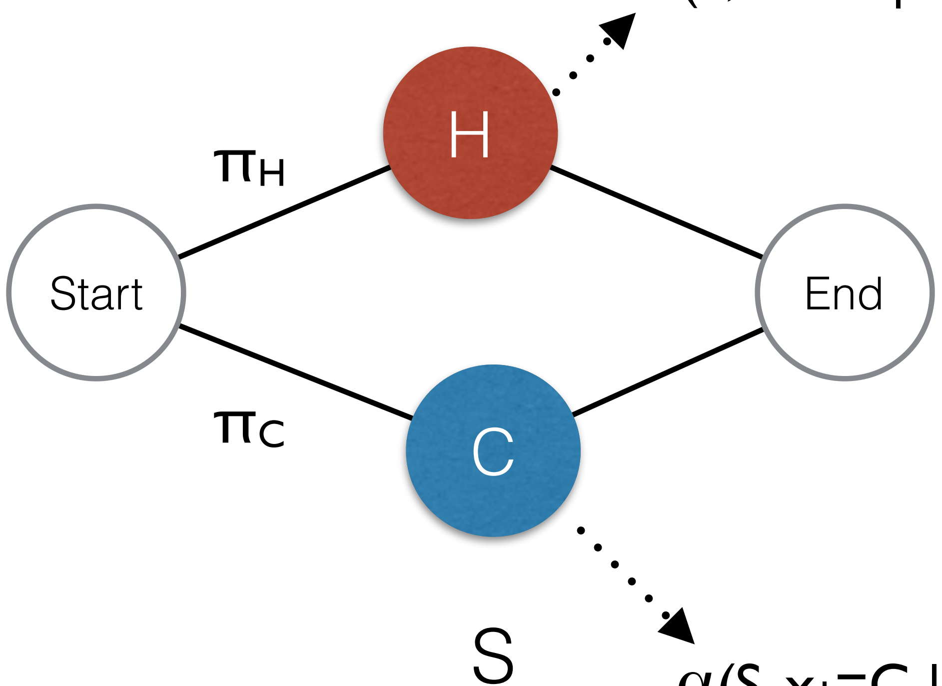


Hidden Markov Models

- Problem 1: To find $P(O|\lambda)$ in a feasible time we use the forward algorithm $O(S)$

t=0 t=1

$$\alpha(S, x_1 = H \mid \lambda) = \alpha_1(H) = \pi_H e_1(S)$$



$$P(O|\lambda) = \alpha_1(H) + \alpha_1(C)$$

$$\alpha(S, x_1 = C \mid \lambda) = \alpha_1(C) = \pi_C e_1(S)$$

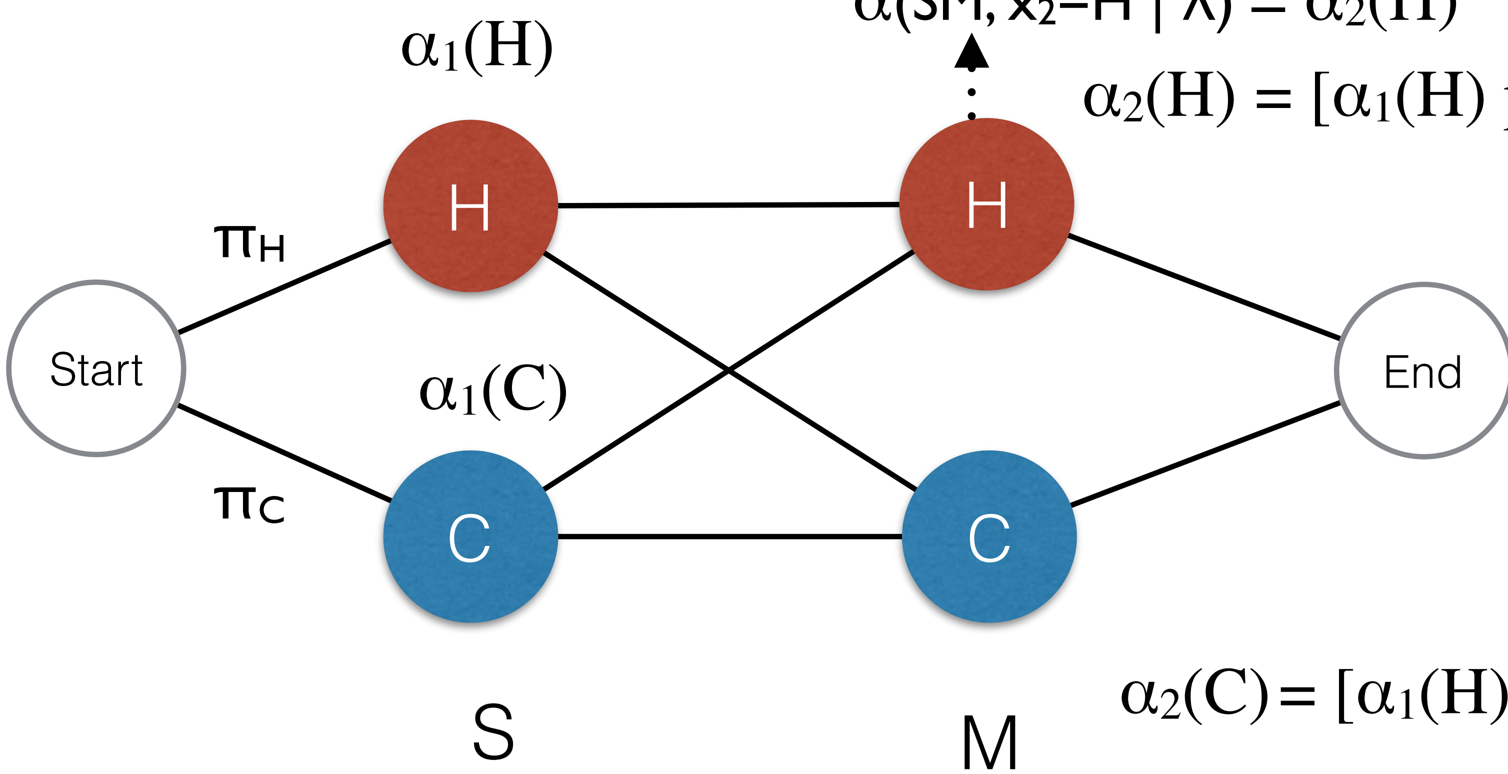
Hidden Markov Models

- Problem 1: To find $P(O|\lambda)$ in a feasible time we use the forward algorithm $O(S,M)$

t=0 t=1 t=2

$\alpha(SM, x_2=H \mid \lambda) = \alpha_2(H)$

$\alpha_2(H) = [\alpha_1(H)\, p_{HH} + \alpha_1(C)\, p_{CH}]\, e_H(M)$



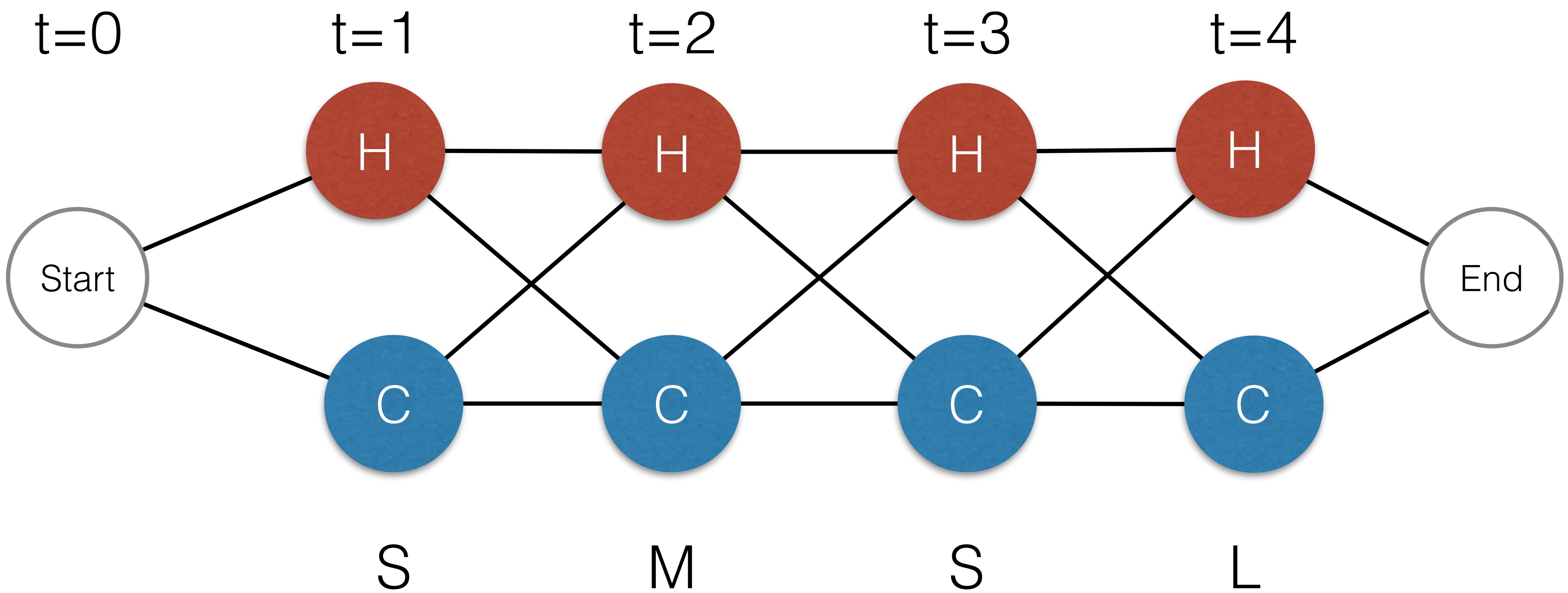
$P(O|\lambda) = \alpha_2(H) + \alpha_2(C)$

$\alpha_2(C) = [\alpha_1(H)\, p_{HC} + \alpha_1(C)\, p_{CC}]\, e_C(M)$

$$\boxed{\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}}$$

$O(S,M,S,L)$

$$P = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad E = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{ccc} S & M & L \end{array} \\ \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \qquad \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ [0.6 \; 0.4] \end{array}$$



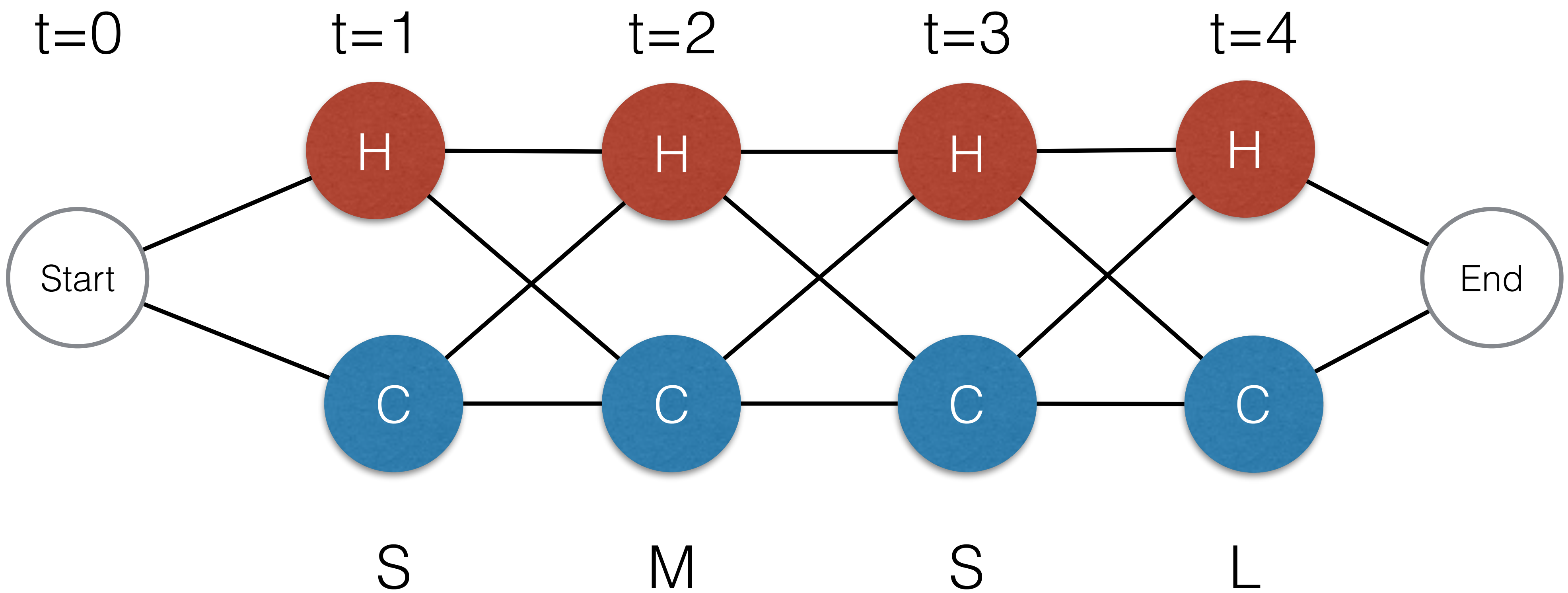
Initialization: $\alpha_0(x_i) = 1$

$$\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1$				
C	$\alpha_0(C)=1$				

$O(S,M,S,L)$

$$P= \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad E= \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{ccc} S & M & L \end{array} \\ \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \qquad \begin{array}{cc} H & C \end{array} \; [0.6 \; 0.4]$$



Initialization: $\alpha_0(x_i) = 1$

$$\boxed{\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1$	$\longrightarrow \alpha_1(H) = 0.06$			
C	$\alpha_0(C)=1$				

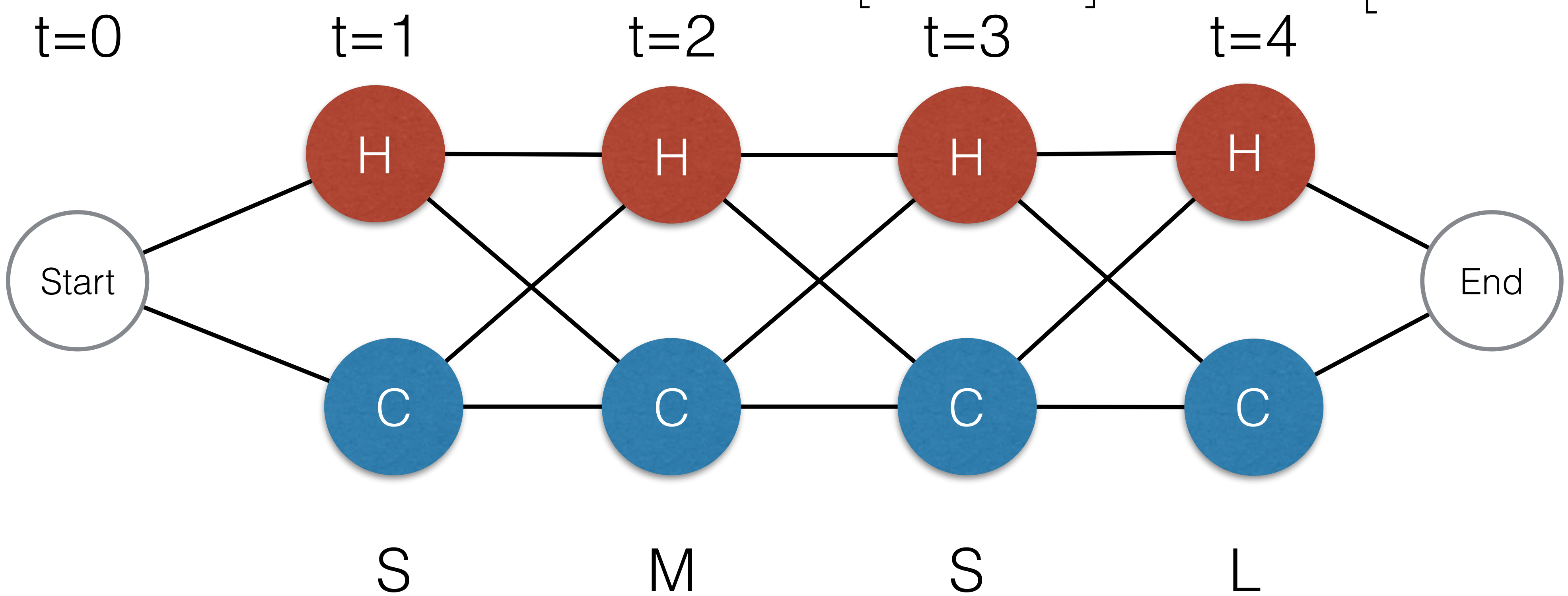
$$\alpha_1(H) = e_H(a_1)\, [\alpha_0(H)\ \ \pi_H]$$

$$\alpha_1(H) = 0.1\ (1)\ 0.6 = 0.06$$

$O(S,M,S,L)$

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

H C
[0.6 0.4]



$$\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1 \longrightarrow$	$\alpha_1(H) = 0.06$			
C	$\alpha_0(C)=1 \longrightarrow$	$\alpha_1(C) = 0.28$			

$$\alpha_1(C) = e_C(a_1)\,[\alpha_0(C)\ \ \pi_C]$$

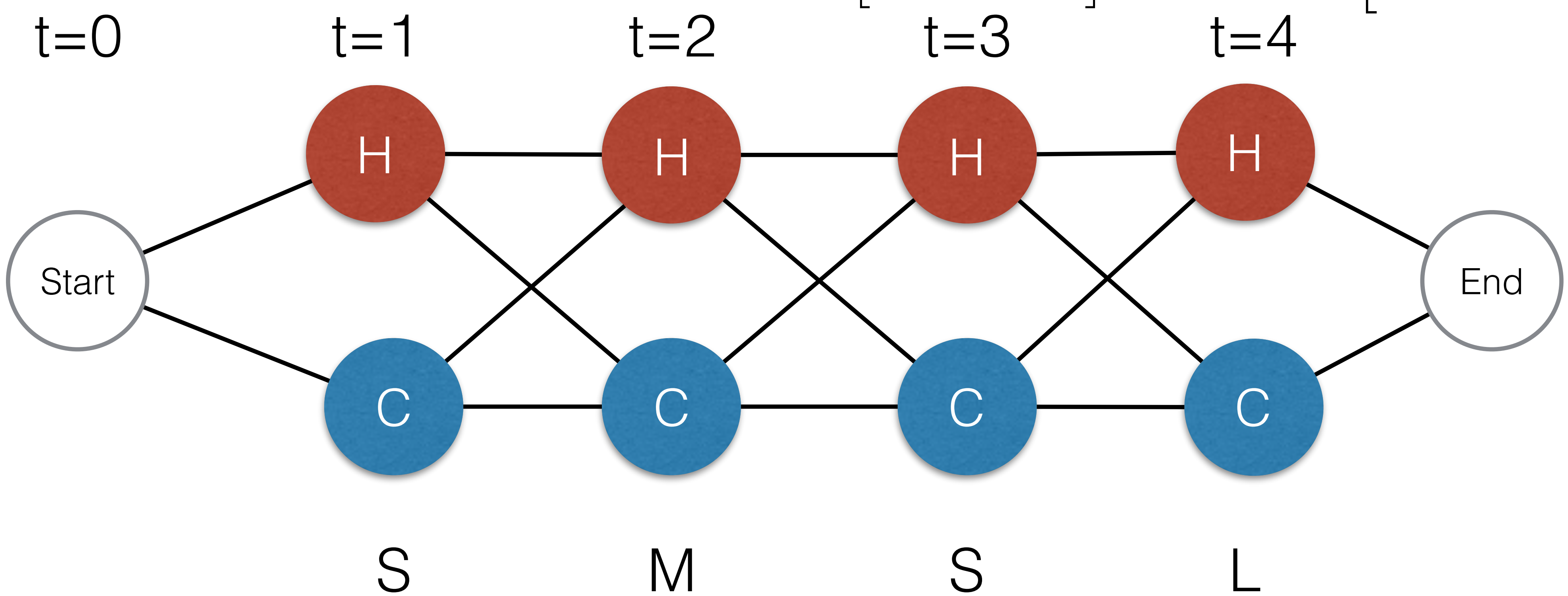
$$\alpha_1(C) = 0.7\ (1)\ 0.4 = 0.28$$

O(S,M,S,L)

$$P = \begin{array}{c|cc} & H & C \\ \hline H & 0.7 & 0.3 \\ C & 0.4 & 0.6 \end{array}$$

$$E = \begin{array}{c|ccc} & S & M & L \\ \hline H & 0.1 & 0.4 & 0.5 \\ C & 0.7 & 0.2 & 0.1 \end{array}$$

H	C
[0.6	0.4]



$$\boxed{\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1 \longrightarrow$	$\alpha_1(H) = 0.06 \longrightarrow$	$\alpha_2(H) = 0.0616$		
C	$\alpha_0(C)=1 \longrightarrow$	$\alpha_1(C) = 0.28 \nearrow$			

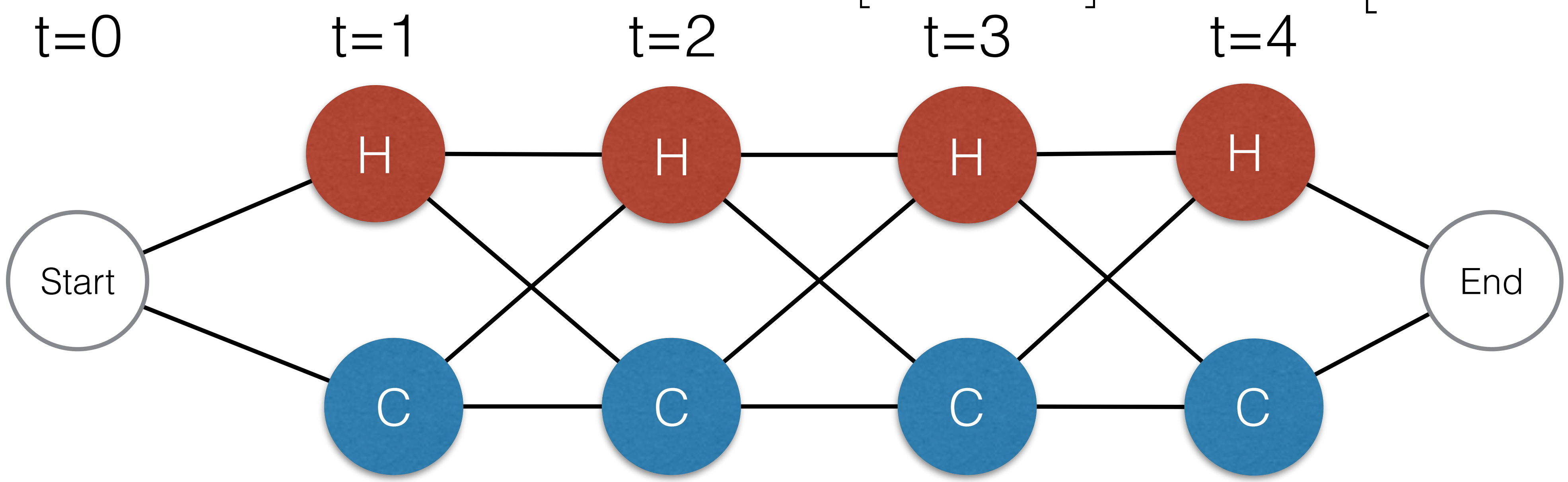
$$\alpha_2(H) = e_H(M)\,[\alpha_1(H)\, p_{HH} + \alpha_1(C)\, p_{CH}]$$

$$\alpha_2(H) = 0.4\,[0.06\,(0.7) + 0.28\,(0.4)] =$$

O(S,M,S,L)

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

H C
[0.6 0.4]



S M S L

Initialization:

$\alpha_0(x_i) = 1$

$$\boxed{\alpha_i(x_i) = e_i(a_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1$	$\alpha_1(H) = 0.06$	$\alpha_2(H) = 0.0616$		
C	$\alpha_0(C)=1$	$\alpha_1(C) = 0.28$	$\alpha_2(C) = 0.03$		

$\alpha_2(C) = e_C(M)\, [\alpha_1(H)\, p_{HC} + \alpha_1(C)\, p_{CC}]$

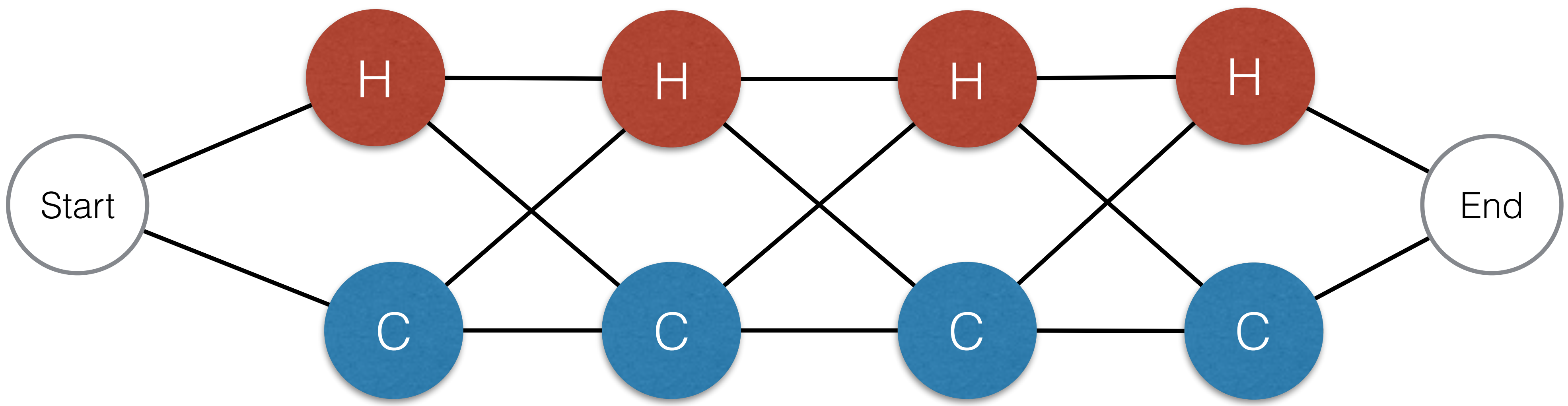
$\alpha_2(H) = 0.2\, [0.06\, (0.3) + 0.28\, (0.6)] = 0.03$

O(S,M,S,L)

$$P = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{cc} H & C \\ \left[\begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array}\right] \end{array} \qquad E = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{ccc} S & M & L \end{array} \\ \left[\begin{array}{ccc} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{array}\right] \end{array}$$

H C
[0.6 0.4]

t=0 t=1 t=2 t=3 t=4



S M S L

Initialization:

$$\alpha_0(x_i) = 1$$

$$\boxed{\alpha_i(x_i) = e_i(O_i) \sum_{j=1}^{N} \alpha_{i-1}(x_j)\, p_{ji}}$$

	t=0	t=1	t=2	t=3	t=4
H	$\alpha_0(H)=1$	$\alpha_1(H) = 0.06$	$\alpha_2(H) = 0.0616$	$\alpha_3(H) = 0.0054$	$\alpha_4(H) = 0.0069$
C	$\alpha_0(C)=1$	$\alpha_1(C) = 0.28$	$\alpha_2(C) = 0.03$	$\alpha_3(C) = 0.025$	$\alpha_4(C) = 0.00831$

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_i(x_T) \qquad P(O|\lambda) = 0.01521$$

The forward algorithm only requires about N^2T multiplications, as opposed to more than $2TN^T$ for the naıve approach

Hidden Markov Models

- There are three fundamental problems that we can solve using HMMs
  - Problem 1: ✔

    Given the model $\lambda = (P, E, \pi)$ and a sequence of observations O, find $P(O \mid \lambda)$.

    Here, we want to determine the likelihood of the observed sequence O, given the model.

  - Problem 2:

    Given $\lambda = (P, E, \pi)$ and an observation sequence O, find an optimal state sequence for the underlying Markov process.

  - Problem 3:

    Given an observation sequence O and the dimensions N and M, find the model $\lambda = (P,E,\pi)$ that maximizes the probability of O.

    This can be viewed as training a model to best fit the observed data

Hidden Markov Models

- Problem 2:

Given λ = (P, E, π) and an observation sequence O, find an optimal state sequence for the underlying Markov process.

➡Viterbi algoritm

Input: An observed sequence O

Output: A hidden path X maximising P(O,X | λ)

Initialization: $v_0(x_i) = 1$

for i = 1 to T do

$$v_i(x_i) = e_i(a_i)\ \max_j(v_{i-1}(x_j)p_{ji})$$
$$ptr_i(x_j) = \mathrm{argmax}_j(v_{i-1}(x_j)p_{ji})$$

return $P(O, X) = \max_j(v_j(T))$

return $i_t^* = \mathrm{argmax}_j(v_j(T))$

Hidden Markov Models

- Problem 2:

Given λ = (P, E, π) and an observation sequence O, find an optimal state sequence for the underlying Markov process.

➡Viterbi algoritm

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \quad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

H C
[0.6 0.4]

Input: An observed sequence O

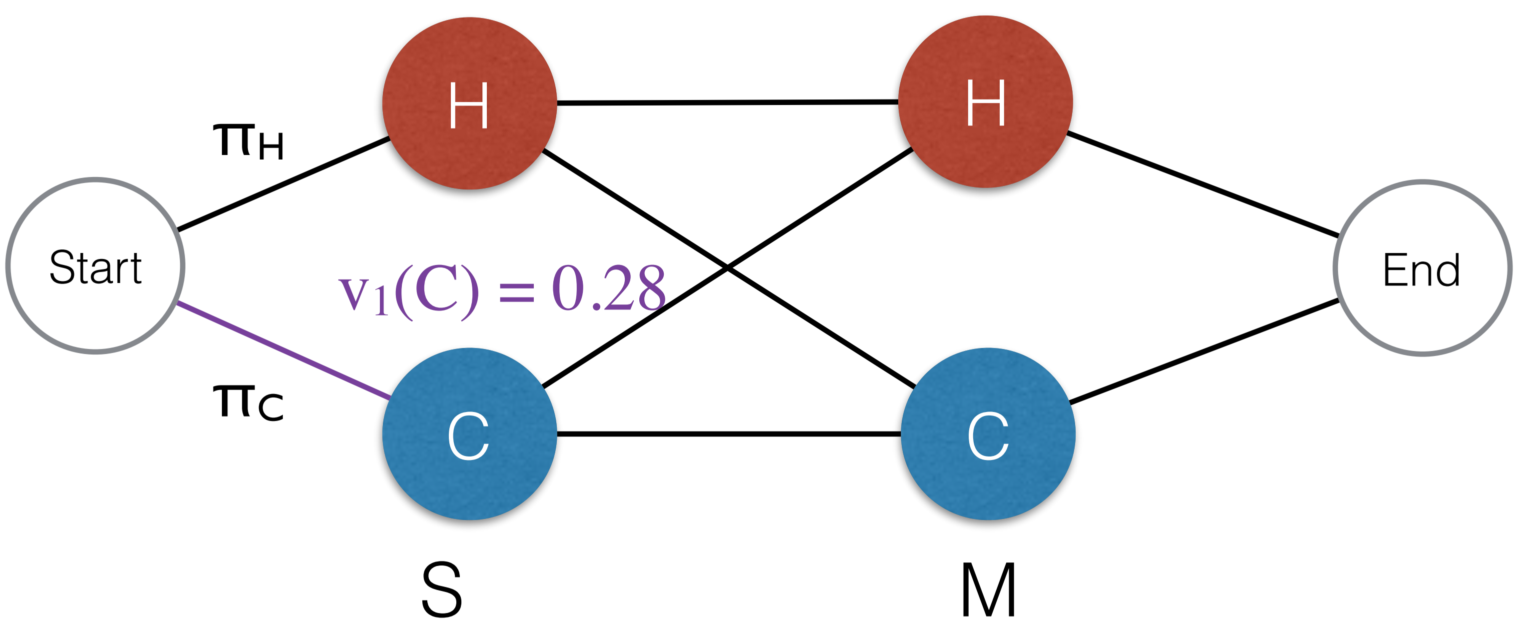
Output: A hidden path X maximising P(O,X | λ)

O(S,M)

t=0 t=1 t=2

$v_1(H) = 0.06$

$v_1(H) = \pi_H \, e_H(S)$



Hidden Markov Models

- Problem 2:

Given $\lambda = (P, E, \pi)$ and an observation sequence O, find an optimal state sequence for the underlying Markov process.

➡Viterbi algoritm

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \quad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \quad \begin{array}{cc} H & C \\ \end{array} \; [0.6 \; 0.4]$$

Input: An observed sequence O

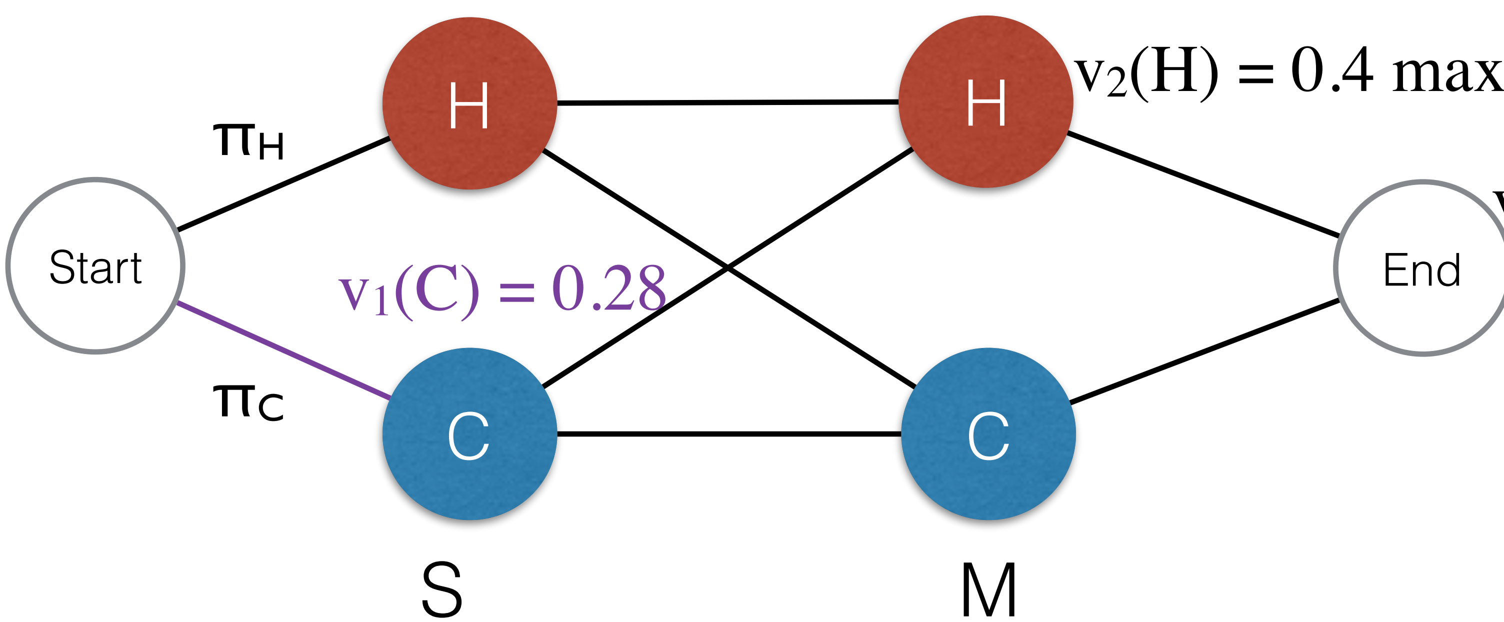
Output: A hidden path X maximising $P(O,X \mid \lambda)$

O(S,M)

t=0 t=1 t=2

$v_1(H) = 0.06$

$v_2(H) \quad = e_H(M) \max[v_1(H)\, p_{HH}, v_1(C)\, p_{CH}]$



$v_1(C) = 0.28$

$v_2(H) = 0.4 \max[0.06\,(0.7), 0.28\,(0.4)] = 0.03$

$v_2(H) = 0.4 \max[0.042, 0.112] = 0.0448$

Hidden Markov Models

- Problem 2:

Given λ = (P, E, π) and an observation sequence O, find an optimal state sequence for the underlying Markov process.

➡Viterbi algoritm

$$P = \begin{array}{c|cc} & H & C \\ \hline H & 0.7 & 0.3 \\ C & 0.4 & 0.6 \end{array} \qquad E = \begin{array}{c|ccc} & S & M & L \\ \hline H & 0.1 & 0.4 & 0.5 \\ C & 0.7 & 0.2 & 0.1 \end{array} \qquad \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

Input: An observed sequence O

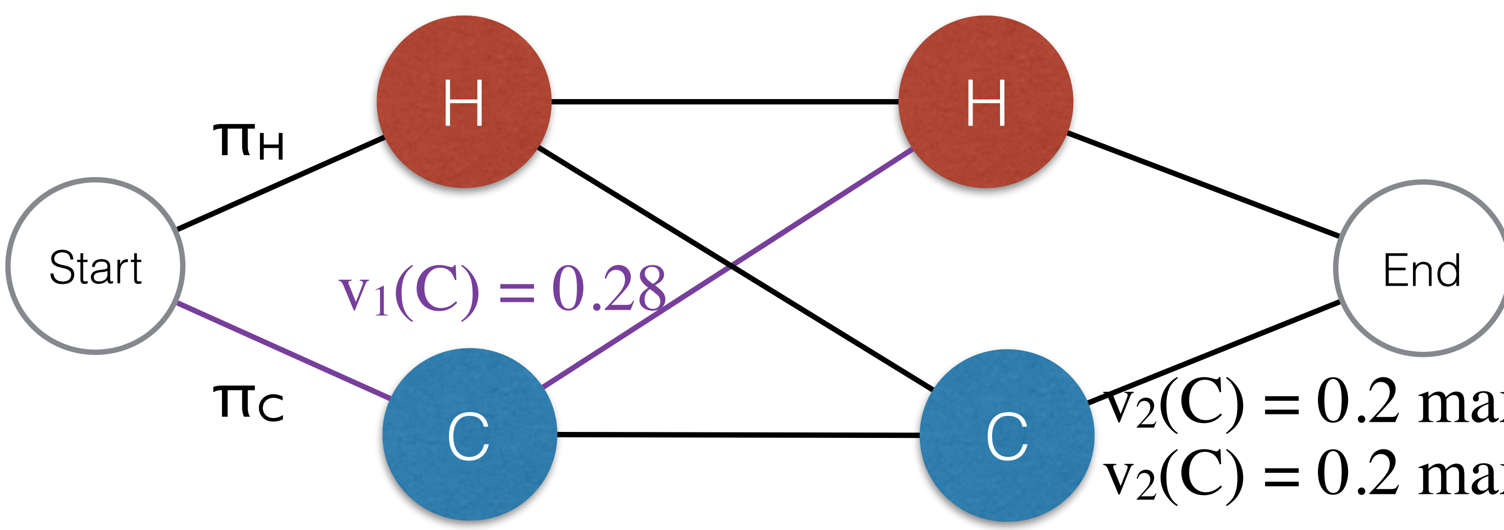
Output: A hidden path X maximising P(O,X | λ)

O(S,M)

t=0 t=1 t=2

$v_1(H) = 0.06$

$v_2(H) = 0.4 \max[0.042, 0.112] = 0.0448$



$v_2(C) = 0.2 \max[0.06\ (0.3), 0.28\ (0.6)] =$

$v_2(C) = 0.2 \max[0.018, 0.168] = 0.0036$

S M

Hidden Markov Models

- Problem 2:

  Given λ = (P, E, π) and an observation sequence O, find an optimal state sequence for the underlying Markov process.

  ➡Viterbi algoritm

$$P = \begin{array}{cc} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \quad E = \begin{array}{cc} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \quad \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

  Input: An observed sequence O

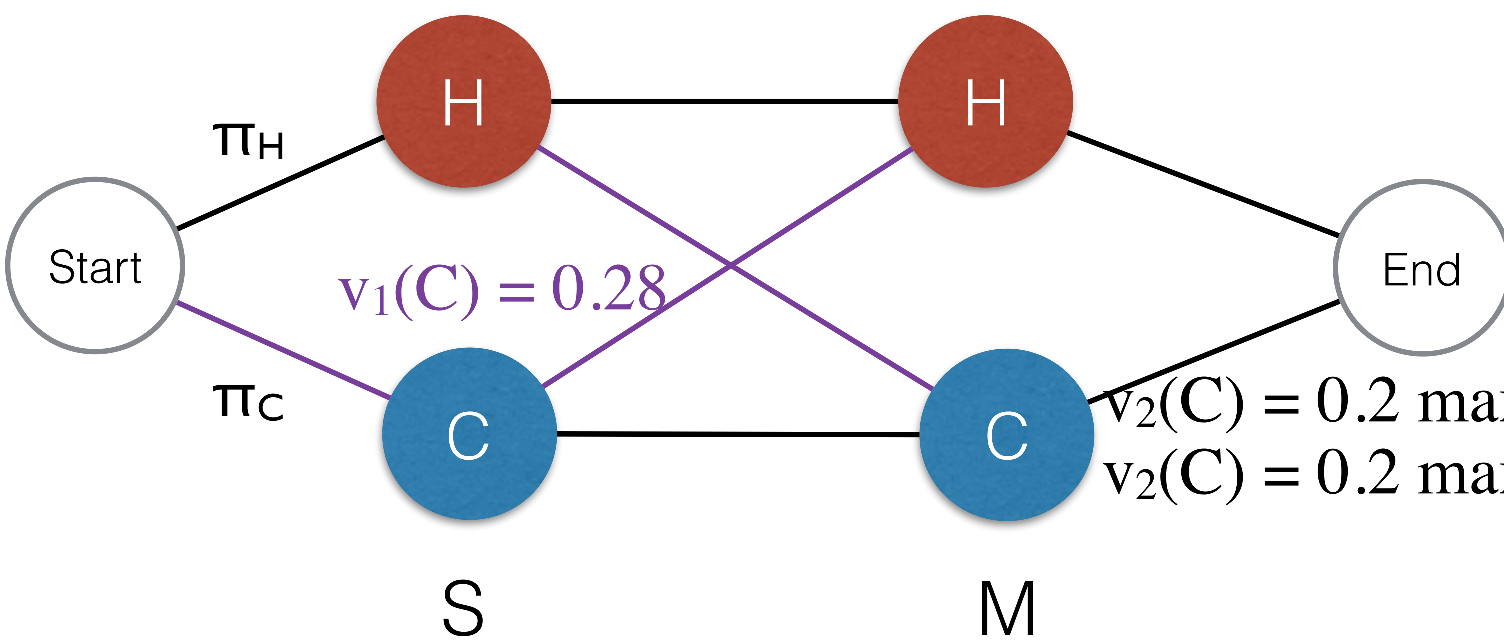
  Output: A hidden path X maximising P(O,X | λ)

O (S,M)

t=0 t=1 t=2

$v_1(H) = 0.06$

$v_2(H) = 0.4 \max[0.042, 0.112] = 0.0448$



$v_2(C) = 0.2 \max[0.06\ (0.3), 0.28\ (0.6)] =$

$v_2(C) = 0.2 \max[0.018, 0.168] = 0.0036$

Hidden Markov Models

- Problem 2:

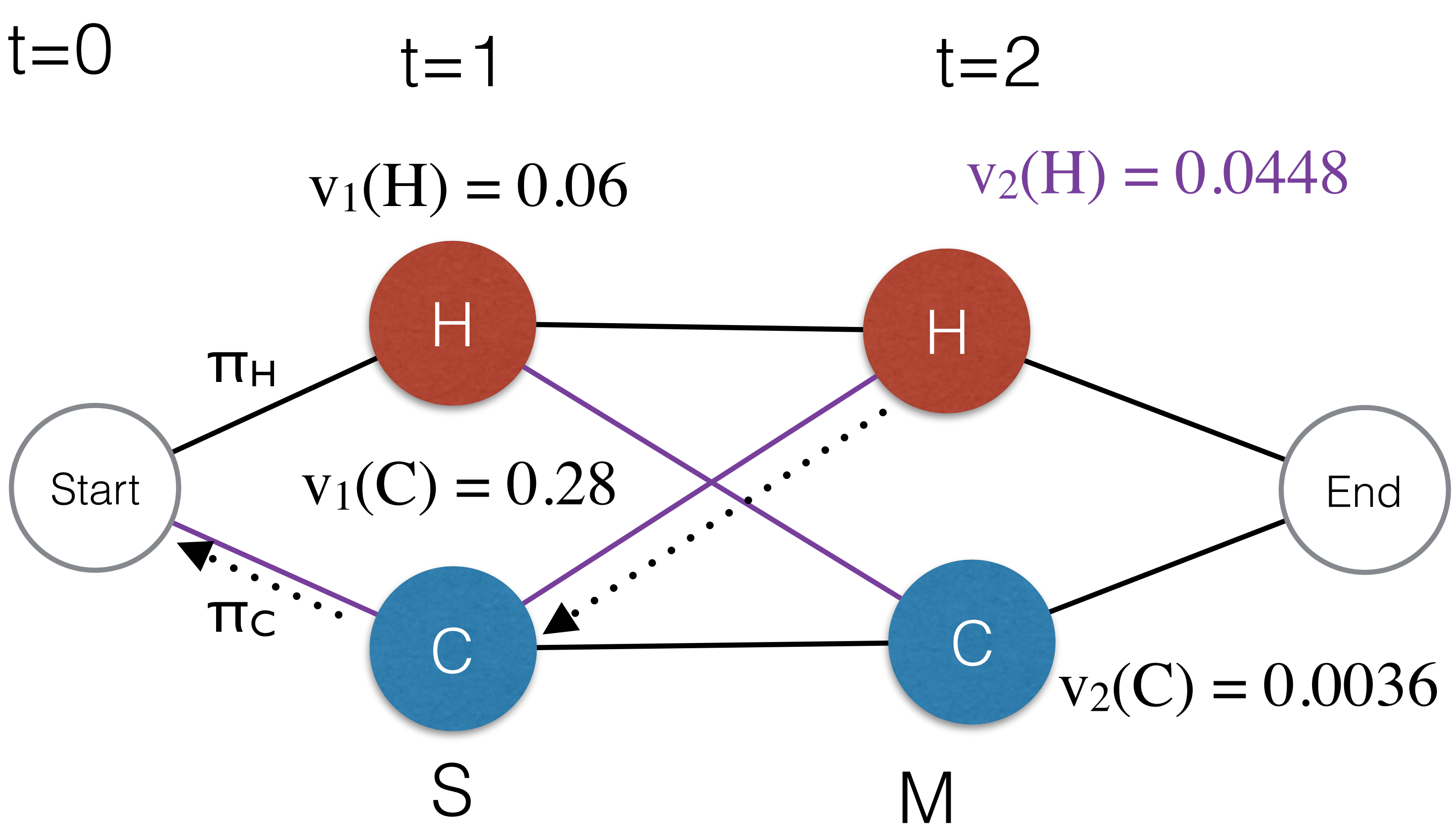
Given $\lambda = (P, E, \pi)$ and an observation sequence O, find an optimal state sequence for the underlying Markov process.

➡Viterbi algoritm

$$P = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \quad E = \begin{array}{c c} & \begin{array}{ccc} S & M & L \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \quad \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

Input: An observed sequence O

Output: A hidden path X maximising $P(O, X \mid \lambda)$



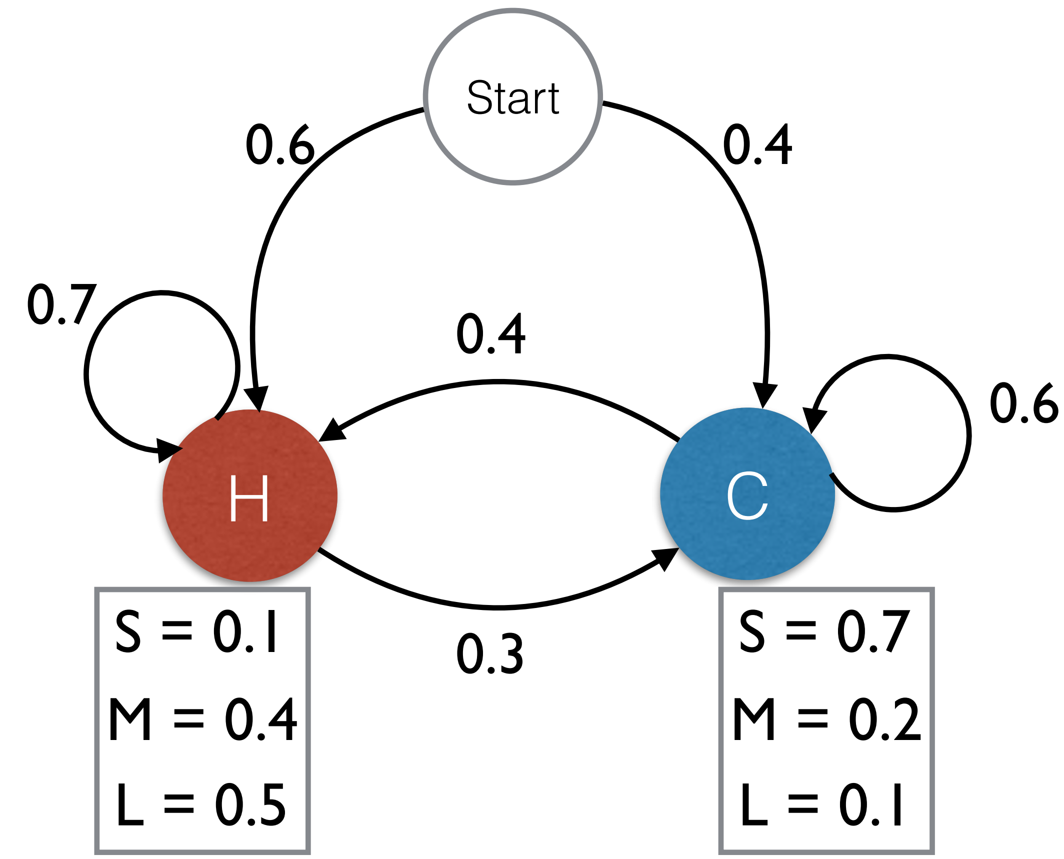
O(S,M)

X(C,H)

$P(O, X \mid \lambda) = 0.0448$

Hidden Markov Models

- Simulating trajectories



$X = x_1, x_2, x_3, \ldots, x_T$

$O = a_1, a_2, a_3, \ldots, a_T$

$T = 4$

X_1

$\downarrow$

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

$\downarrow$

H

S = 0.1
M = 0.4
L = 0.5

$\downarrow$

L

We take a random number y [0,1]

if y <0.6 then x_1 = H

else x_1 = C

We take a random number w [0,1]

if w <0.1 then O_1 = S

else if w >=0.1 and w < 0.5 then O_1 = M

else O_1 = L

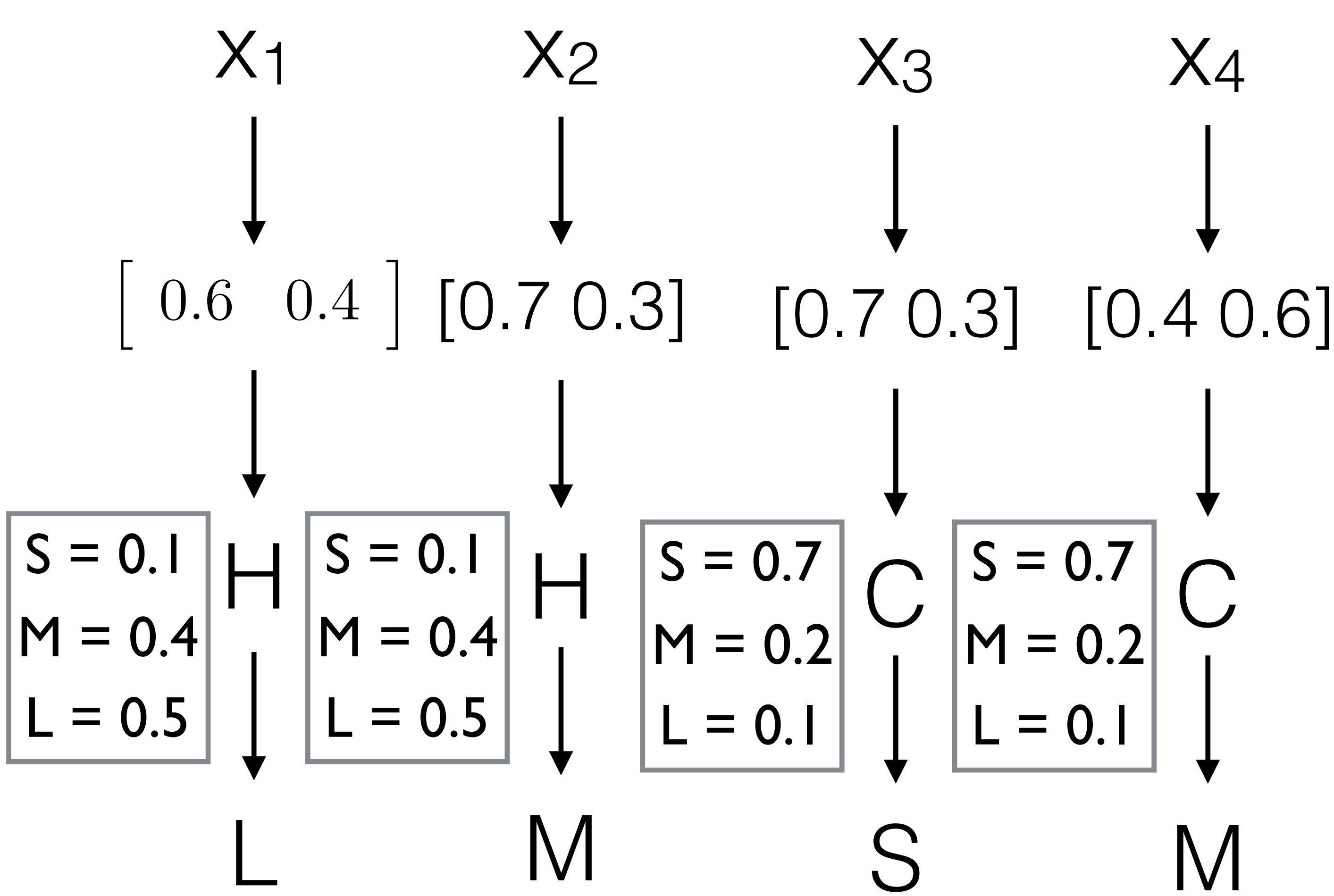
Hidden Markov Models

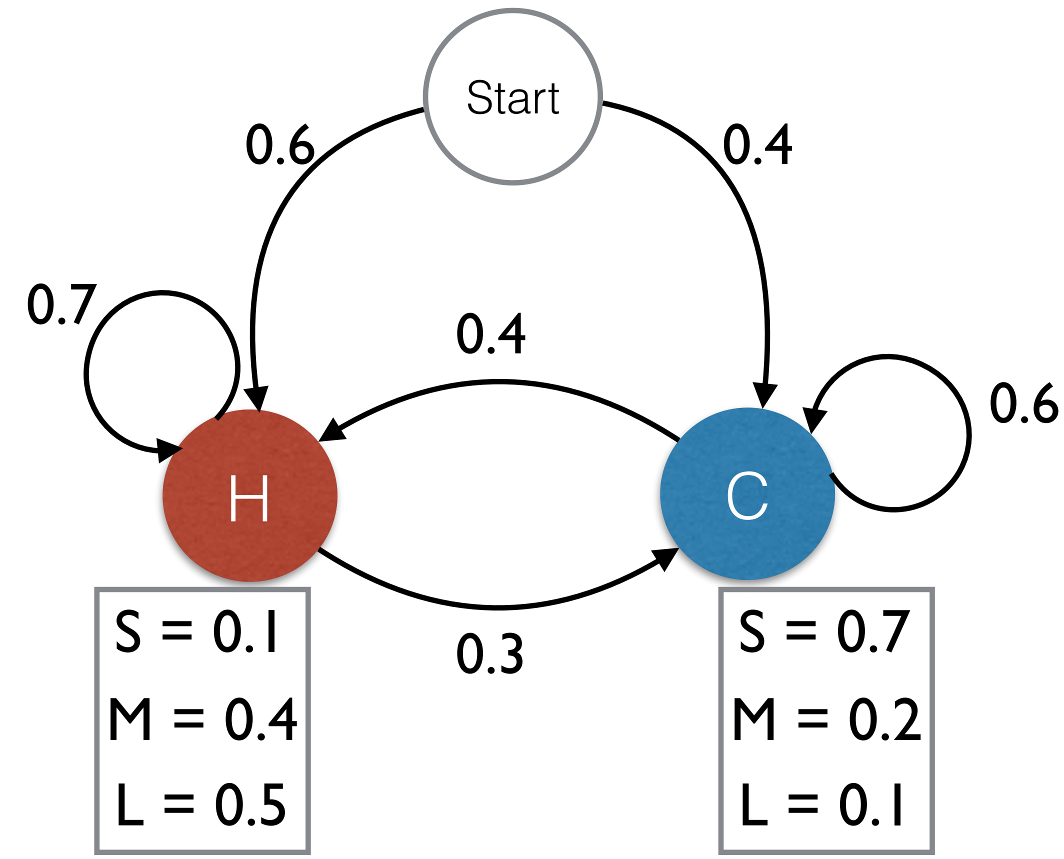
- Simulating trajectories

$$X = x_1, x_2, x_3, \ldots, x_T$$

$$O = a_1, a_2, a_3, \ldots, a_T$$

$$T = 4$$





Hidden Markov Models

- There are three fundamental problems that we can solve using HMMs
  - Problem 1: ✔

    Given the model $\lambda = (P, E, \pi)$ and a sequence of observations O, find $P(O \mid \lambda)$.

    Here, we want to determine the likelihood of the observed sequence O, given the model.

  - Problem 2: ✔

    Given $\lambda = (P, E, \pi)$ and an observation sequence O, find an optimal state sequence for the underlying Markov process.

  - Problem 3:

    Given an observation sequence O and the dimensions N and M, find the model $\lambda = (P,E,\pi)$ that maximizes the probability of O.

    This can be viewed as training a model to best fit the observed data

Hidden Markov Models

- Problem 3:

  Given an observation sequence O and the dimensions N and M, find the model λ = (P,E,π) that maximizes the probability of O.

  This can be viewed as training a model to best fit the observed data

  If we have O and X we can learn P and E by counting

$$O = a_1, a_2, a_3, \ldots, a_T$$

$$X = x_1, x_2, x_3, \ldots, x_T$$

$$p_{ij} = \frac{N_{ij}}{\sum_{k=1} N_{ik}} \qquad e_i(a) = \frac{M_{ia}}{\sum_{b=1} M_{ib}}$$

X= HHCCHCCHHCHCH

O= LMSMMLSSLSMLM

$$p_{HH} = \frac{N_{HH}}{N_{HH} + N_{HC}}$$

$$p_{HH} = \frac{2}{2 + 4} = 1/3$$

Hidden Markov Models

- Problem 3:

  Given an observation sequence O and the dimensions N and M, find the model λ = (P,E,π) that maximizes the probability of O.

  This can be viewed as training a model to best fit the observed data

  If we have O and X we can learn P and E by counting

$$O = a_1, a_2, a_3, \ldots, a_T$$

$$X = x_1, x_2, x_3, \ldots, x_T$$

X= HHCCHCCHHCHCH

O= LMSMMLSSLSMLM

$$p_{ij} = \frac{N_{ij}}{\sum_{k=1} N_{ik}} \qquad e_i(a) = \frac{M_{ia}}{\sum_{b=1} M_{ib}}$$

$$p_{HH} = 1/3 \qquad p_{HC} = 2/3$$

$$p_{CC} = 1/3 \qquad p_{CH} = 2/3$$

$$e_H(S) = \frac{M_{HS}}{M_{HS} + M_{HM} + M_{HL}}$$

Hidden Markov Models

- Problem 3:

  Given an observation sequence O and the dimensions N and M, find the model $\lambda = (P, E, \pi)$ that maximizes the probability of O.

  This can be viewed as training a model to best fit the observed data

  If we have O and X we can learn P and E by counting

$$O = a_1, a_2, a_3, \ldots, a_T$$

$$X = x_1, x_2, x_3, \ldots, x_T$$

X= HHCCHCCHHCHCH

O= LMSMMLSSLSMLM

$$p_{ij} = \frac{N_{ij}}{\sum_{k=1} N_{ik}} \qquad e_i(a) = \frac{M_{ia}}{\sum_{b=1} M_{ib}}$$

$$p_{HH} = 0.2 \qquad p_{HC} = 0.4$$

$$p_{CC} = 0.2 \qquad p_{CH} = 0.4$$

$$e_H(S) \frac{M_{HS}}{M_{HS} + M_{HM} + M_{HL}} = \frac{1}{1 + 4 + 2} = 0.14$$

Hidden Markov Models

- Problem 3:

  Given an observation sequence O and the dimensions N and M, find the model $\lambda = (P, E, \pi)$ that maximizes the probability of O.

  This can be viewed as training a model to best fit the observed data

  If we have O and X we can learn P and E by counting

$$O = a_1, a_2, a_3, \ldots, a_T$$

$$X = x_1, x_2, x_3, \ldots, x_T$$

We have O and we can estimate X by Viterbi so we can learn P and E

Viterbi training

Hidden Markov Models

- Problem 3: Viterbi training

Algorithm: Viterbi training

input: A sequence of Observations O and the number of iteration I

output: P and E matrix

Initiate P and E with random values

for(i = 1 to I) do

$i_t^* = \text{Viterbi}(O, \pi, P, E)$

$[P, E] = \text{count}(O, i_t^*)$

How large is I ?

Hidden Markov Models

- Problem 3: Viterbi training

Algorithm: Viterbi training

input: A sequence of Observations O and the number of iteration I

output: P and E matrix

Initiate P, E and L with random values

do

$[L_{new}, i_t^*] = \text{Viterbi}(O, \pi, P, E)$

$[P, E] = \text{count}(O, i_t^*)$

if (L_{new} != L) then L=L_{new}

else stop = true

while stop == false