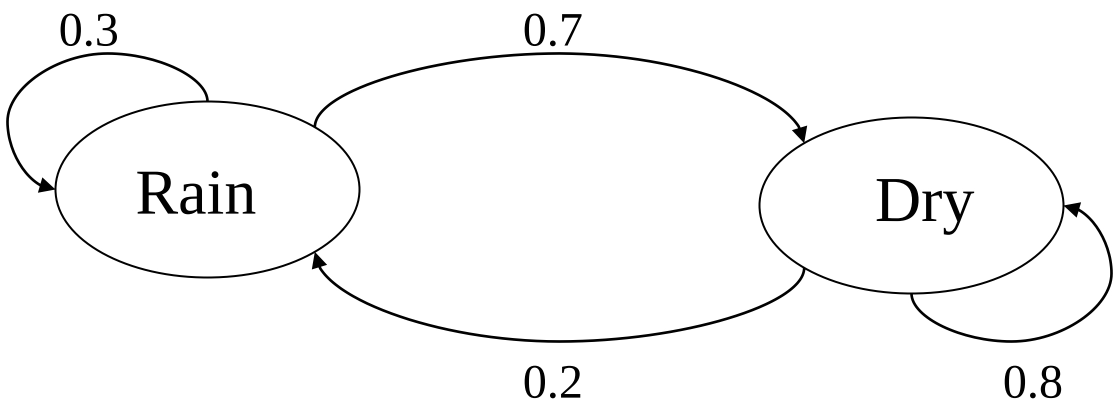
Hidden Markov Models (Examples)

Donghai Guan

Nanjing University of Aeronautics and Astronautics (NUAA)

Example of Markov Model



- Two states : ‘Rain’ and ‘Dry’.
- Transition probabilities:

P(‘Rain’|‘Rain’)=0.3 , P(‘Dry’|‘Rain’)=0.7 ,

P(‘Rain’|‘Dry’)=0.2, P(‘Dry’|‘Dry’)=0.8

- Initial probabilities: say P(‘Rain’)=0.4 , P(‘Dry’)=0.6 .

Calculation of sequence probability

$$P(S_1 = Dry, S_2 = Dry, S_3 = Rain)$$
$$= \underline{P(S_1 = Dry)P(S_2 = Dry \mid S_1 = Dry)P(S_3 = Rain \mid S_2 = Dry)}$$
$$= 0.6 \times 0.8 \times 0.2$$

Markov Model is just a Bayesian Network;

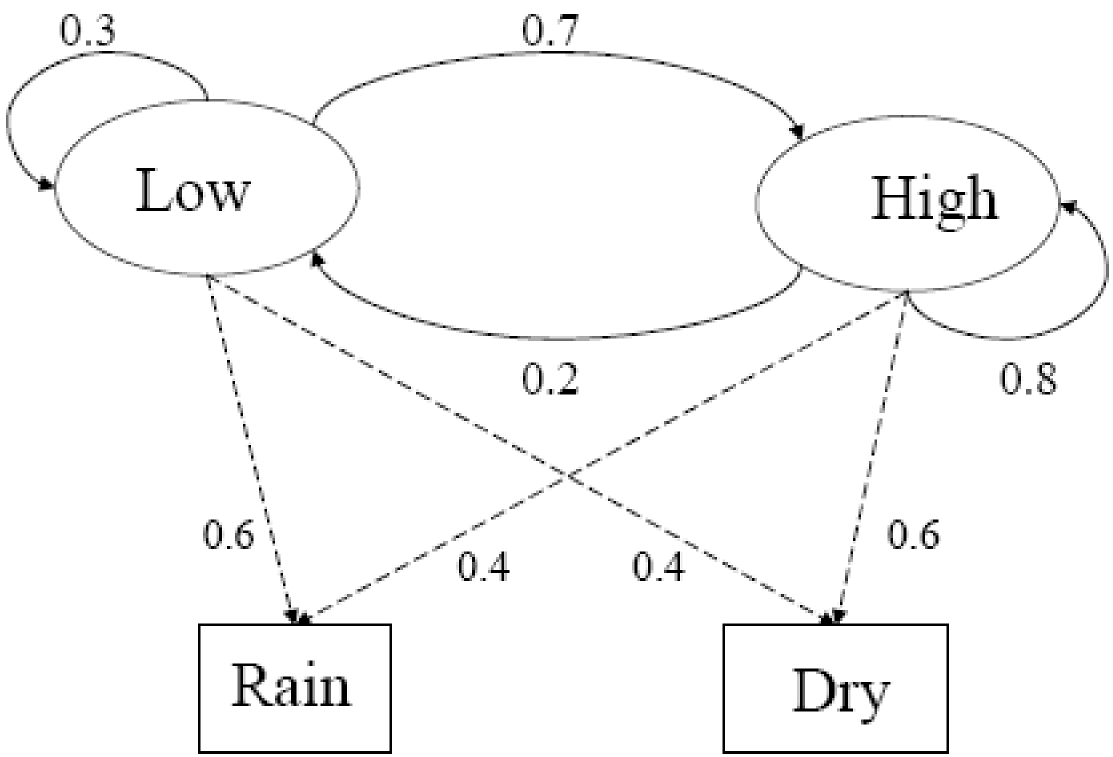
In this network, P(S1,S2,S3)=P(S1)P(S2|S1)P(S3|S2)

Hidden Markov Models

- Based on Markov Models
- Differences include
  - State becomes “Hidden”
  - The state information is not available, instead of, there are some “Observations” with are correlated with “Hidden” State

Hidden Markov Model

Example 1



Transition probability matrix:

	Low	High
Low	0.3	0.7
High	0.2	0.8

Emission probability matrix:

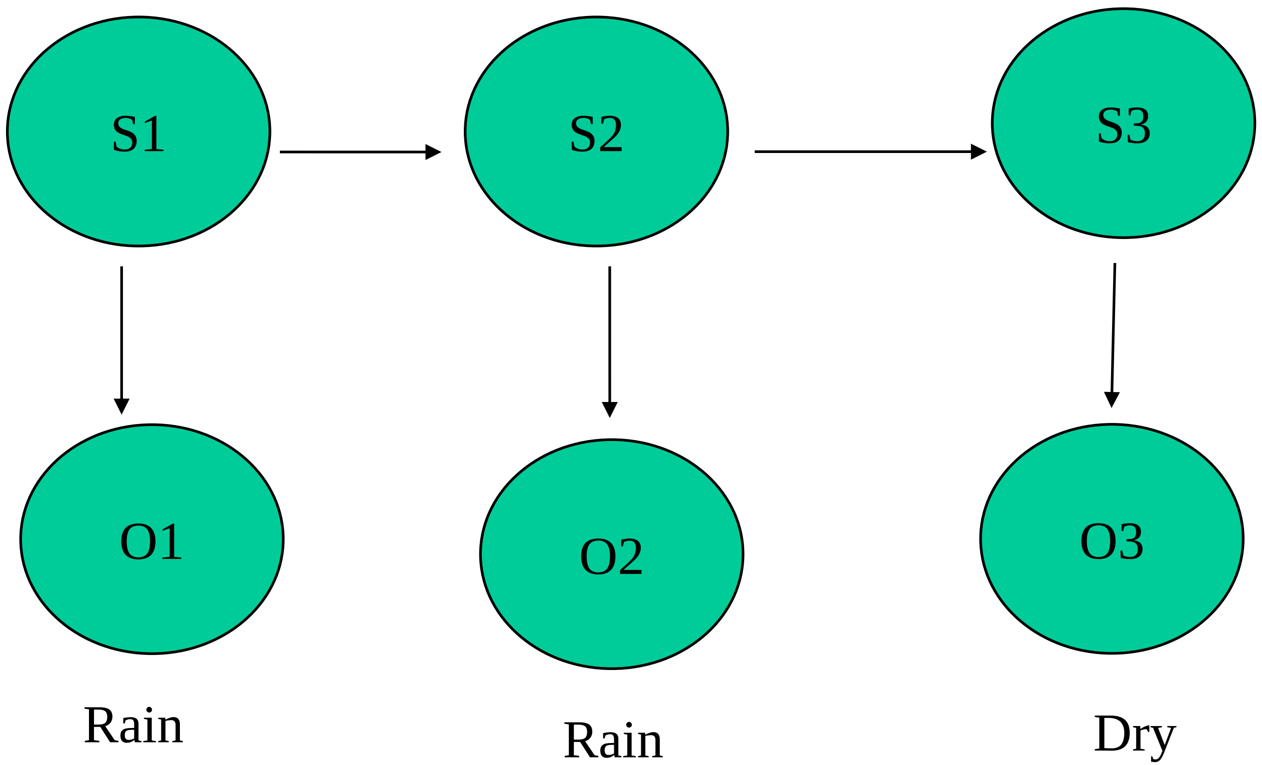
	Rain	Dry
Low	0.6	0.4
High	0.4	0.3

Two states : 'Low' and 'High' atmospheric pressure.
Two observations : 'Rain' and 'Dry'.
Initial probabilities: P('Low')=0.4 , P('High')=0.6 .

In Markov Model,
Emission Prob. is not used

Problems can be solved using HMM

1) Calculation of observation sequence probability



$$P(O1 = Rain, O2 = Rain, O3 = Dry)$$

$$P(O1 = Rain, O2 = Rain, O3 = Dry)$$

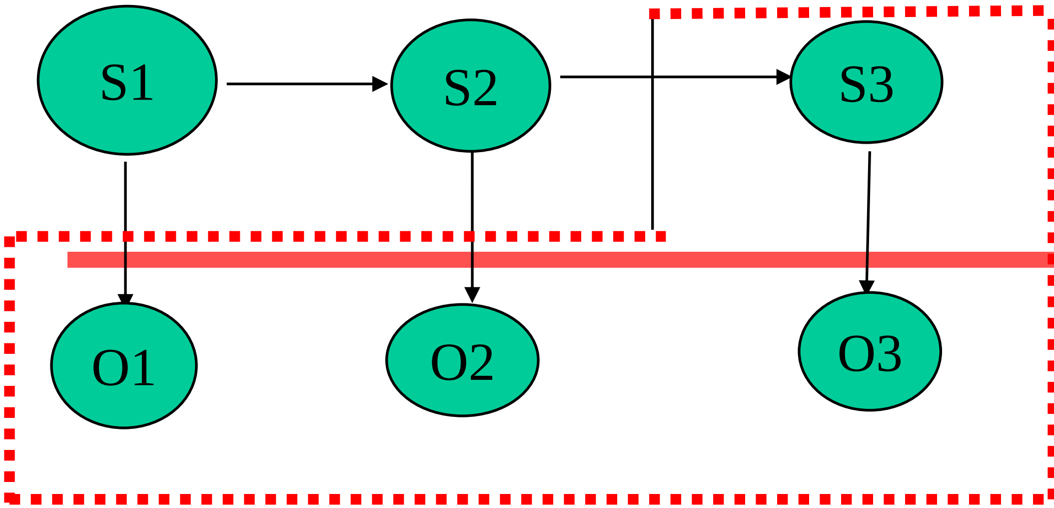
$$= \sum_{S3=\{low,high\}} P(O1 = Rain, O2 = Rain, O3 = Dry, S3)$$

$$P(O1 = Rain, O2 = Rain, O3 = Dry, S3 = Low) = \alpha_3^{Low}$$

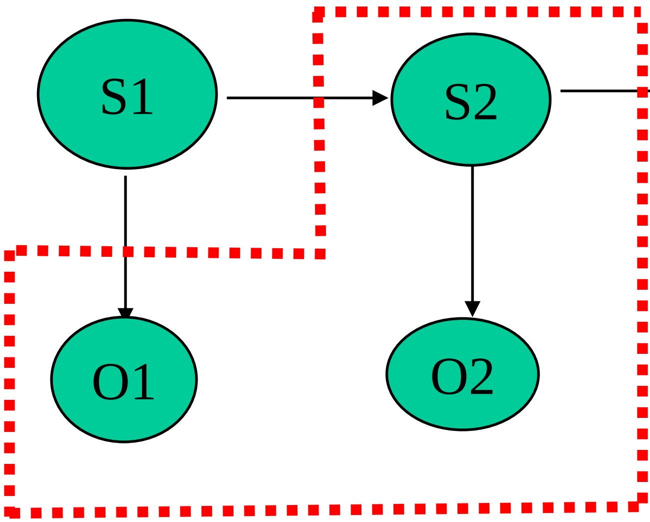
$$P(O1 = Rain, O2 = Rain, O3 = Dry, S3 = High) = \alpha_3^{High}$$

$$P(O1 = Rain, O2 = Rain, O3 = Dry) = \alpha_3^{Low} + \alpha_3^{High} = \sum_{i\in\{Low,High\}} \alpha_3^{i}$$

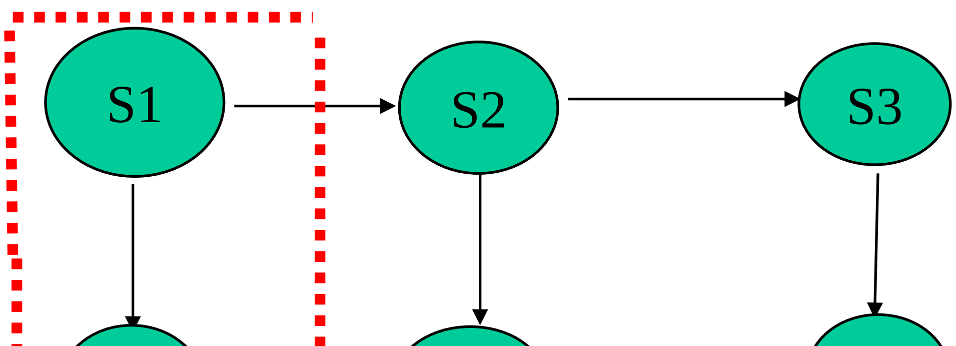
* Now the problem is how to calculate α_3^{Low} & α_3^{High}



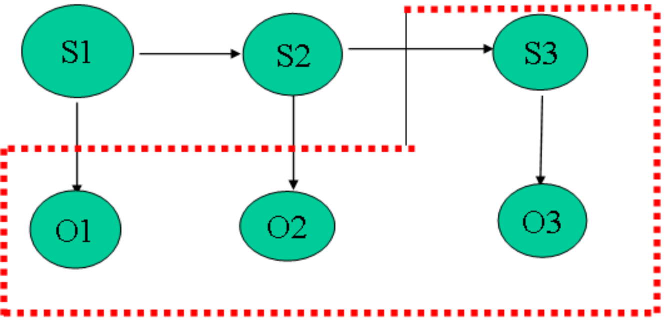
$$\alpha_3^i = P(O1, O2, O3, S3 = i)$$



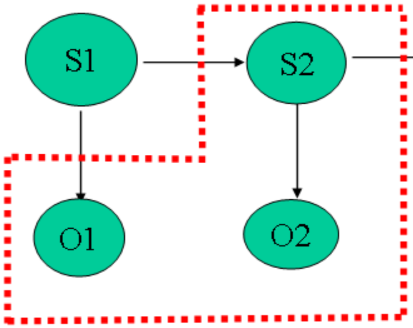
$$\alpha_2^i = P(O1, O2, S2 = i)$$



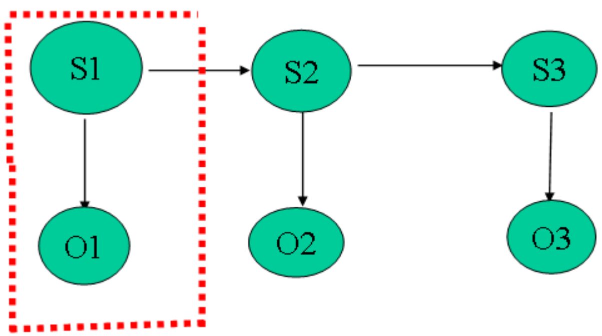
$$\alpha_1^i = P(O1, S1 = i)$$



$$\alpha_3^i = P(O1, O2, O3, S3 = i)$$



$$\alpha_2^i = P(O1, O2, S2 = i)$$



$$\alpha_1^i = P(O1, S1 = i)$$

Calculation Difficulty

Can we find some relationship between α_3^i & α_2^i & α_1^i

If we can find the relationship, then we can:

1) Calculate α_1^i
2) Calculate α_2^i based on α_1^i
3) Calculate α_3^i based on α_2^i

Recursively!

Can we find the relationship? (Yes)

$$\alpha_3^{Low} = P(O1, O2, O3, S3 = Low)$$

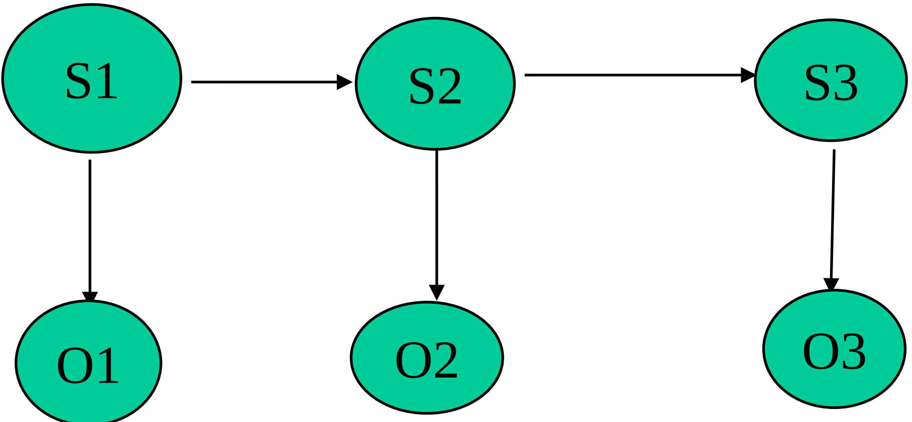
$$= P(O1, O2, O3, S3 = Low, S2 = High) + P(O1, O2, O3, S3 = Low, S2 = Low)$$

$$= \sum_{j=\{Low,High\}} P(O1, O2, O3, S3 = Low, S2 = j)$$

$$= \sum_{j=\{Low,High\}} P(O1, O2, S2 = j) \times \underline{P(O3 \mid S3 = Low, O1, O2, S2 = j)} \times \underline{P(S3 = Low \mid O1, O2, S2 = j)}$$

D-Separation

$$P(O3 \mid S3 = Low) \qquad\qquad P(S3 = Low \mid S2 = j)$$



Forward Algorithm

$$\alpha_3^{Low} = \sum_{j=\{Low,High\}} P(O1, O2, S2 = j) \times P(O3 \mid S3 = Low) \times P(S3 = Low \mid S2 = j)$$

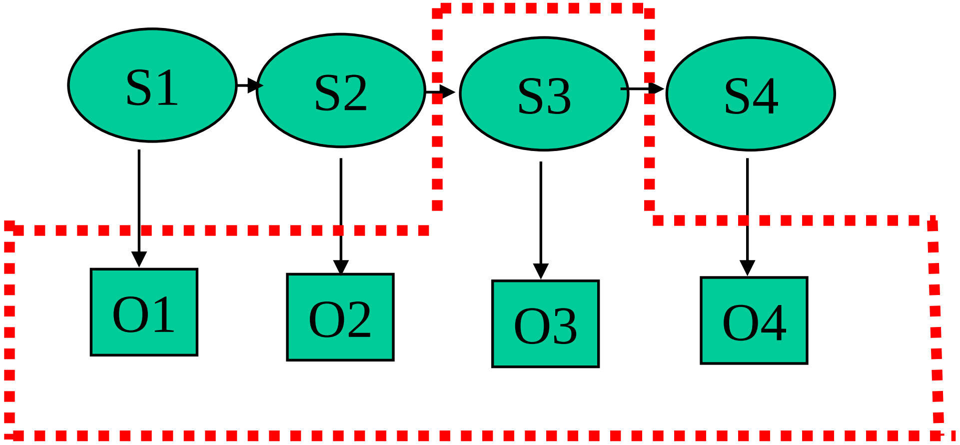
$$= \sum_{j=\{Low,High\}} \alpha_2^j \times P(O3 \mid S3 = Low) \times P(S3 = Low \mid S2 = j)$$

$$= \underbrace{P(O3 \mid S3 = Low)}_{\text{Emission Probability}} \sum_{j=\{Low,High\}} \alpha_2^j \times \underbrace{P(S3 = Low \mid S2 = j)}_{\text{Transition Probability}}$$

Dynamic Programming

Problem can be Solved by HMM

- Decoding problem 1



P(O1,O2,O3,O4,S3=High)=?

$$P(O1, O2, O3, O4, S3 = High) =$$

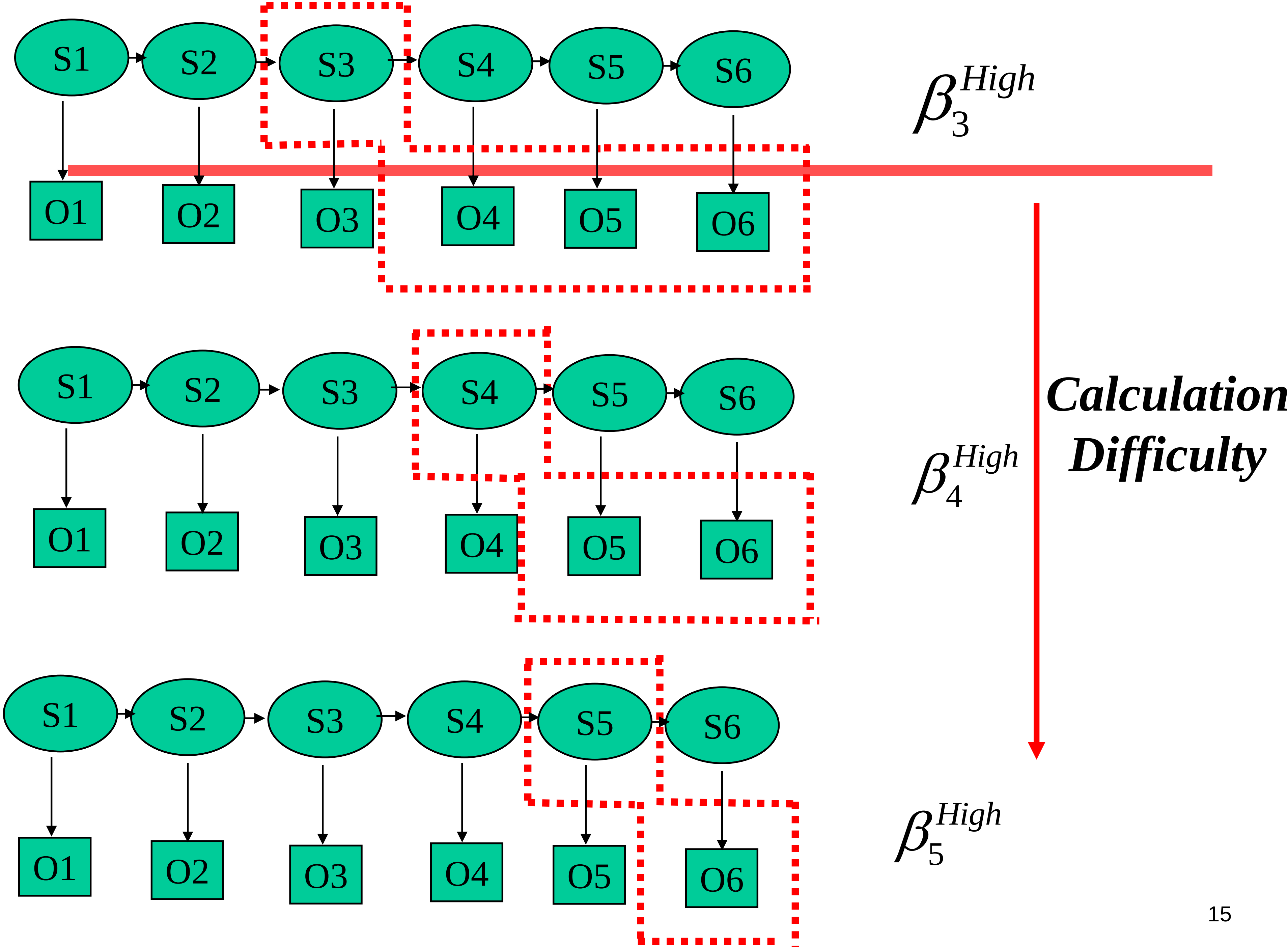
$$P(O1, O2, O3, S3 = High) \times P(O4 \mid O1, O2, O3, S3 = High)$$

$$= P(O1, O2, O3, S3 = High) \times P(O4 \mid S3 = High)$$

$$= \underline{\alpha_3^{High}} \times \underline{\beta_3^{High}}$$

We already know how
to calculate it.

?



S1
S2
S3
S4
S5
S6
O1
O2
O3
O4
O5
O6
β_3^{High}
β_4^{High}
β_5^{High}
Calculation Difficulty

Can we find some relationship among β_3^i & β_4^i & β_5^i & β_6^i

If we can find the relationship, then we can:

1) Calculate $\beta_6^i = 1$
2) Calculate β_5^i based on β_6^i
3) Calculate β_4^i based on β_5^i
4) Calculate β_3^i based on β_4^i

Recursively!

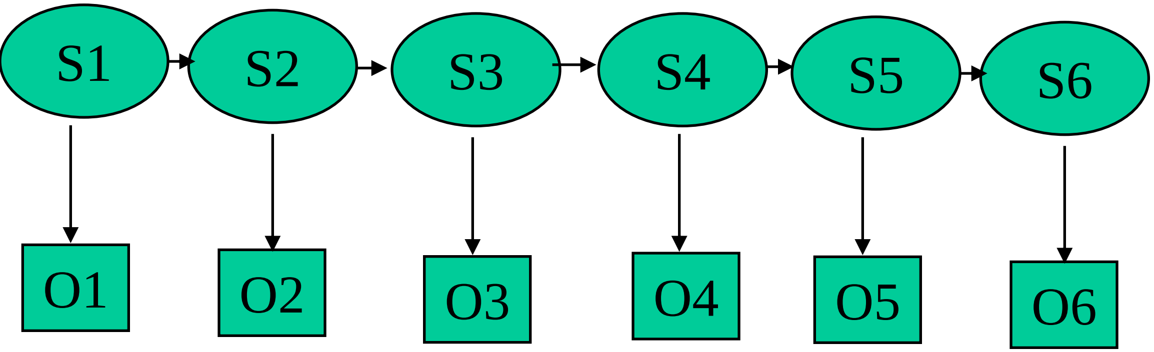
Can we find the relationship? (Yes)

$$\beta_3^{High} = P(O4, O5, O6 \mid S3 = High)$$

$$= \sum_{j=\{Low,High\}} P(O4, O5, O6, S4 = j \mid S3 = High)$$

$$= \sum_{j=\{Low,High\}} \underbrace{P(O5, O6 \mid S4 = j, O4, S3 = High)}_{\beta_4^j} \times \underbrace{P(O4 \mid S4 = j, S3 = High)}_{P(O4 \mid S4 = j)} \times P(S4 = j \mid S3 = High)$$

D-Separation



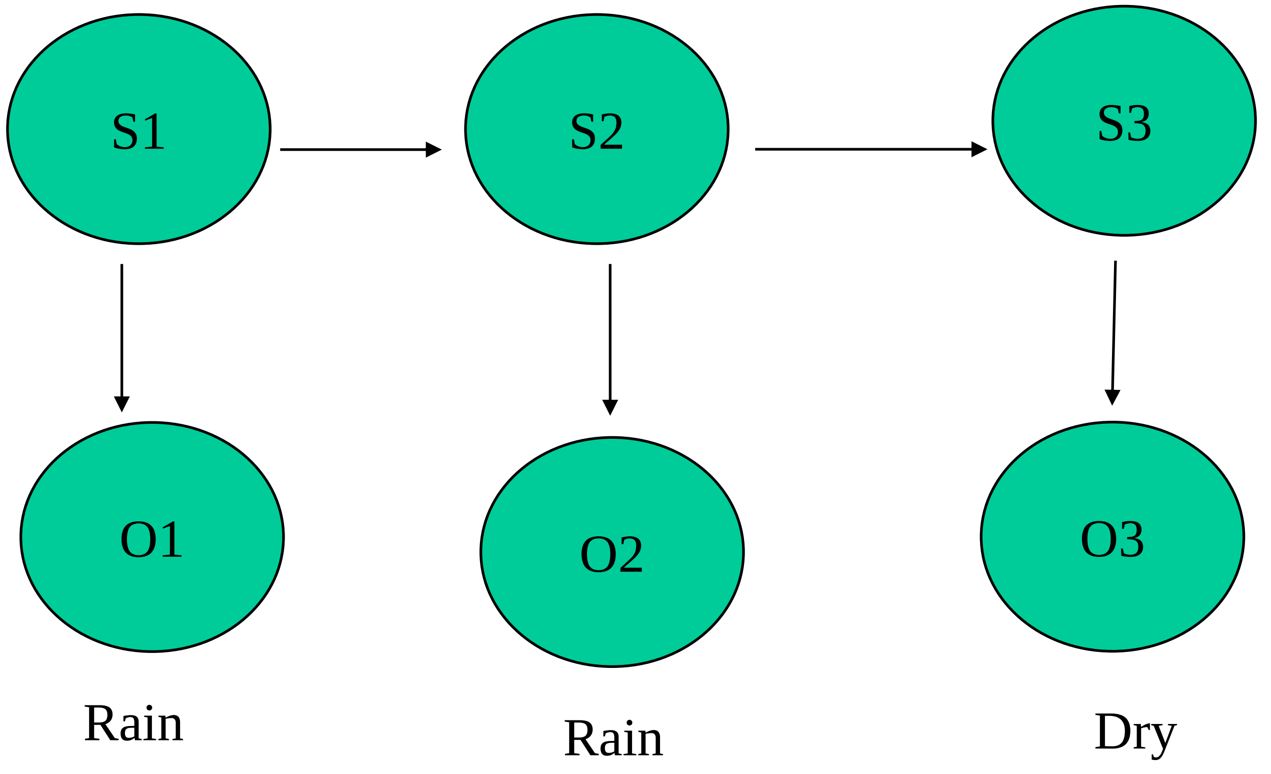
Backward Algorithm

$$\beta_3^{High} = \sum_{j=\{Low,High\}} \beta_4^j \times \underline{P(O4 \mid S4 = j)} \times \underline{P(S4 = j \mid S3 = High)}$$

Emission Prob.

Transition Prob.

Decoding Problem 2



O1,O2,O3 are known, what is the most probable sequence S1,S2,S3

For example: {High, High, Low}, {Low, High, Low}, {Low, High, High}…….

$$\arg\max_{S1,S2,S3} P(S1,S2,S3 \mid O1,O2,O3)$$

$$= \arg\max_{S1,S2,S3} P(S1,S2,S3,O1,O2,O3)$$

$$= \arg\max_{k} \underbrace{\max_{S1,S2} P(S3=k,S1,S2,O1,O2,O3)}_{V_3^k}$$

Probability of most likely sequence of states ending at states S3=k

$$V_3^k$$

$$= \max_{S1,S2} P(S3=k, S2, S1, O1, O2, O3)$$

$$= \max_i \max_{S1} P(S3=k, S2=i, S1, O1, O2, O3)$$

$$= \max_i \max_{S1} P(S2=i, S1, O1, O2) P(S3=k, O3 \mid S2=i, S1, O1, O2)$$

$$= \max_i \max_{S1} P(S2=i, S1, O1, O2) P(O3 \mid S3=k, S2=i, S1, O1, O2) P(S3=k \mid S2=i, S1, O1, O2)$$

D-Separation

$$P(O3 \mid S3=k) \qquad P(S3=k \mid S2=i)$$

$$V_3^k = \max_i V_2^i P(O3 \mid S3=k) P(S3=k \mid S2=i)$$

$$= P(O3 \mid S3=k) \max_i P(S3=k \mid S2=i) V_2^i$$

Viterbi algorithm

$$V_3^k = \max_i V_2^i P(O3 \mid S3 = k) P(S3 = k \mid S2 = i)$$

$$= P(O3 \mid S3 = k) \max_i P(S3 = k \mid S2 = i) V_2^i$$

$$V_2^k = P(O2 \mid S2 = k) \max_i P(S2 = k \mid S1 = i) V_1^i$$

Viterbi algorithm

Can compute V_t^k for all k, t using dynamic programming:

- Initialize: $V_1^k = p(O_1 | S_1 = k) p(S_1 = k)$ for all k

- Iterate: for t = 2, …, T

$$V_t^k = p(O_t|S_t = k) \max_i p(S_t = k|S_{t-1} = i) V_{t-1}^i$$ for all k

- Termination: $$\max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = \max_k V_T^k$$

Traceback: $$S_T^* = \arg\max_k V_T^k$$

$$S_{t-1}^* = \arg\max_i p(S_t^*|S_{t-1} = i) V_{t-1}^i$$

Viterbi algorithm

	H	C	L
H	50%	40%	10%
C	10%	80%	10%
L	10%	60%	30%

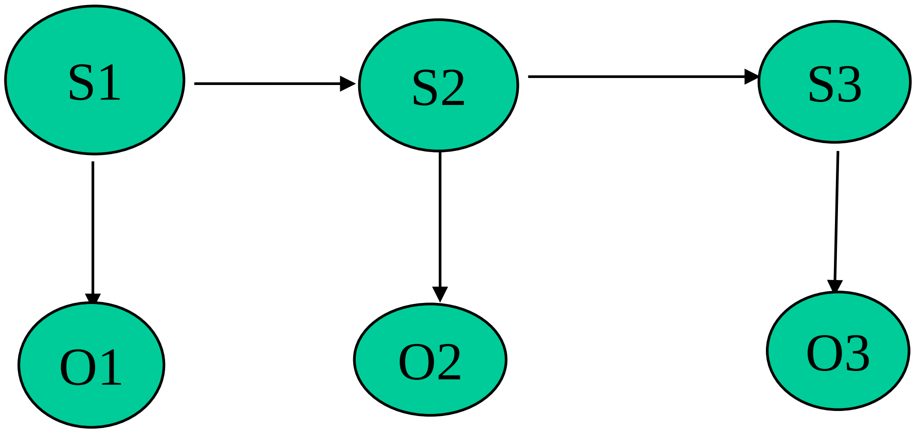
	T	F
H	90%	10%
C	50%	50%
L	10%	90%

H	0.05	0.0025	0.0007
C	0.125	0.0675	0.027
L	0.225	0.0607	0.0164
	F	F	F

The most likely internal state is (L,C,C) with a joint probability of 0.027

Problem can be Solved by HMM

- Learning algorithm



What is the transition probability, P(High|Low)=?, P(Low|High)=?

What is the emission probability, P(Dry|High)=? P(Dry|Low)=?

What is the initial probability, P(S1=High)=?

Baum-Welch Algorithm (EM)

Start with random initialization of parameters

	Low	High
Low	0.6	0.4
High	0.3	0.7

	Rain	Dry
Low	0.7	0.3
High	0.4	0.6

$$P(S1 = Low) = 0.6, P(S1 = High) = 0.4$$

Baum-Welch Algorithm (EM)

- E-Step

$$\gamma_i(t) = p(S_t = i | O, \theta) = \frac{\alpha_t^i \beta_t^i}{\sum_j \alpha_t^j \beta_t^j}$$

Forward-Backward algorithm

$$\xi_{ij}(t) = p(S_{t-1} = i, S_t = j | O, \theta)$$

$$= \frac{p(S_{t-1} = i | O, \theta) p(S_t = j, O_t, \ldots, O_T | S_{t-1} = i, \theta)}{p(O_t, \ldots, O_T | S_{t-1} = i, \theta)}$$

$$= \frac{\gamma_i(t-1)\ p_{ij}\ q_j^{O_t}\ \beta_t^j}{\beta_{t-1}^i}$$

Baum-Welch (EM) Algorithm

- Start with random initialization of parameters
- **E-step**

$$\gamma_i(t) = p(S_t = i|O,\theta)$$

$$\xi_{ij}(t) = p(S_{t-1} = i, S_t = j|O,\theta)$$

$\sum_{t=1}^{T} \gamma_i(t)$ = expected # times in state i

$\sum_{t=1}^{T\text{-}1} \gamma_i(t)$ = expected # transitions from state i

$\sum_{t=1}^{T-1} \xi_{ij}(t)$ = expected # transitions from state i to j

- **M-step**

$$\pi_i = \gamma_i(1)$$

$$p_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$$

$$q_i^k = \frac{\sum_{t=1}^{T} \delta_{O_t=k} \gamma_i(t)}{\sum_{t=1}^{T} \gamma_i(t)}$$