CS344: Introduction to Artificial Intelligence (associated lab: CS386)

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay

Lecture 6-7: Hidden Markov Model

18th and 20th Jan, 2011

Observations leading to why probability is needed

- Many intelligence tasks are sequence labeling tasks
- Tasks carried out in layers
- Within a layer, there are limited windows of information
- This naturally calls for strategies for dealing with uncertainty
- Probability and Markov process give a way

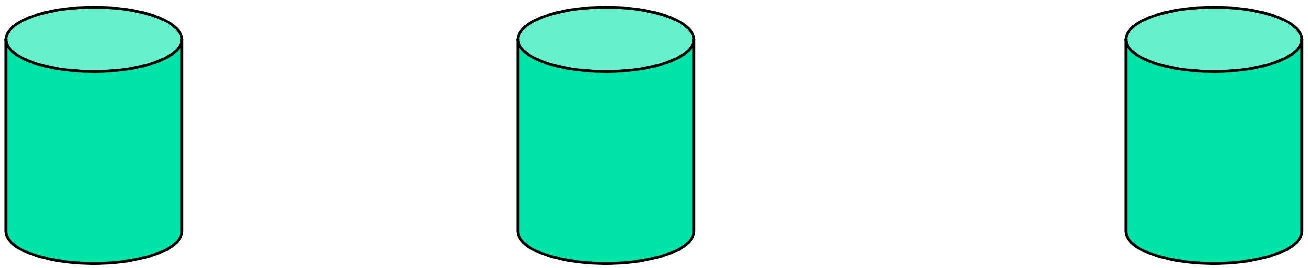
"I went with my friend to the bank to withdraw some money, but was disappointed to find it closed"

POS	Bank (N/V)		closed (V/ adj)
Sense	Bank (financial institution)		withdraw (take away)
Pronoun drop	But	I/friend/money/bank	was disappointed
SCOPE	With	my friend	
Co-referencing	It -> bank		

HMM

A Motivating Example

Colored Ball choosing



Urn 1
# of Red = 30
# of Green = 50
# of Blue = 20

Urn 2
# of Red = 10
# of Green = 40
# of Blue = 50

Urn 3
# of Red =60
# of Green =10
# of Blue = 30

Probability of transition to another Urn after picking a ball:

	U_1	U_2	U_3
U_1	0.1	0.4	0.5
U_2	0.6	0.2	0.2
U_3	0.3	0.4	0.3

Example (contd.)

Given :

	U_1	U_2	U_3
U_1	0.1	0.4	0.5
U_2	0.6	0.2	0.2
U_3	0.3	0.4	0.3

and

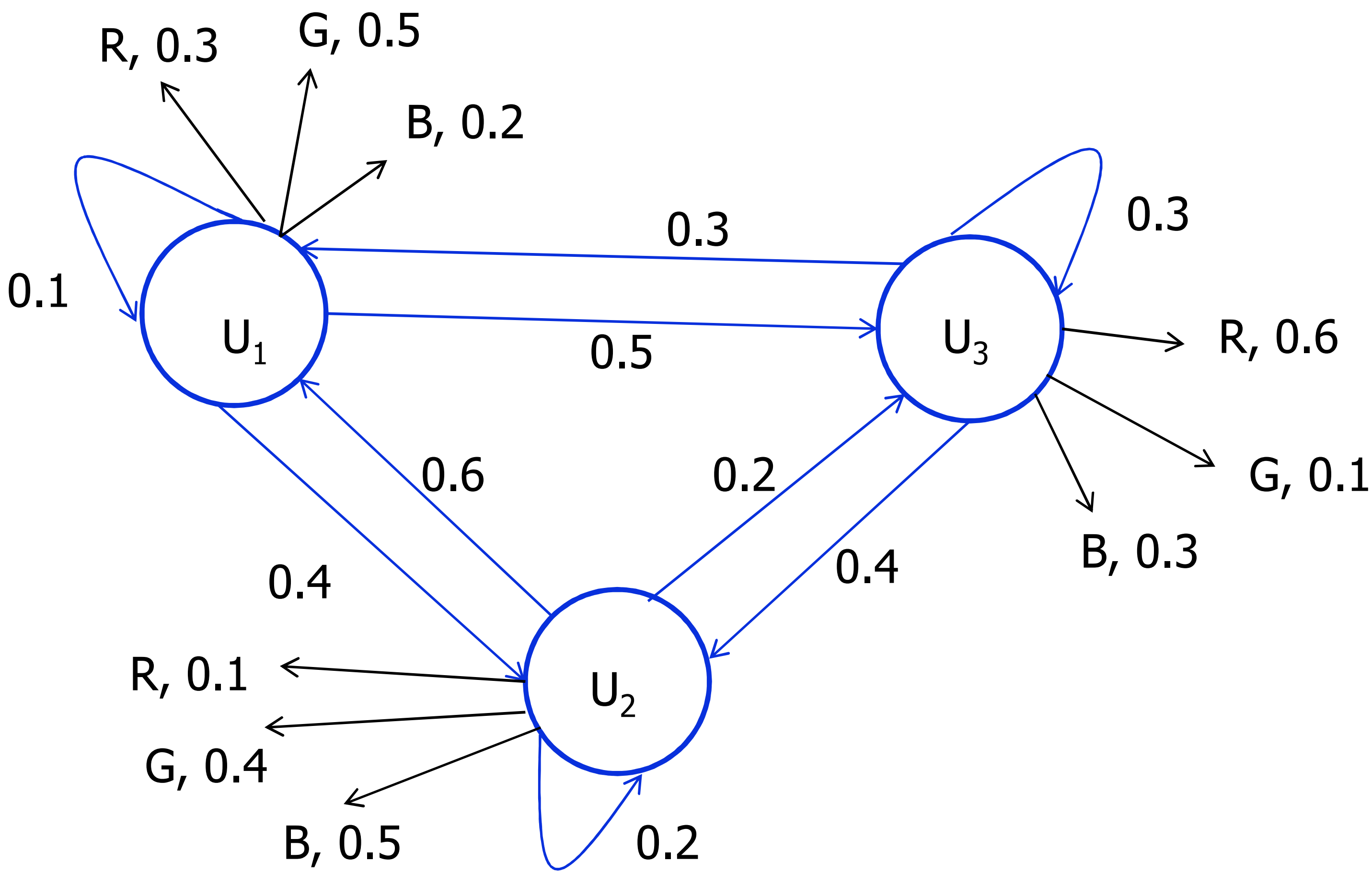
	R	G	B
U_1	0.3	0.5	0.2
U_2	0.1	0.4	0.5
U_3	0.6	0.1	0.3

Observation : RRGGBRGR

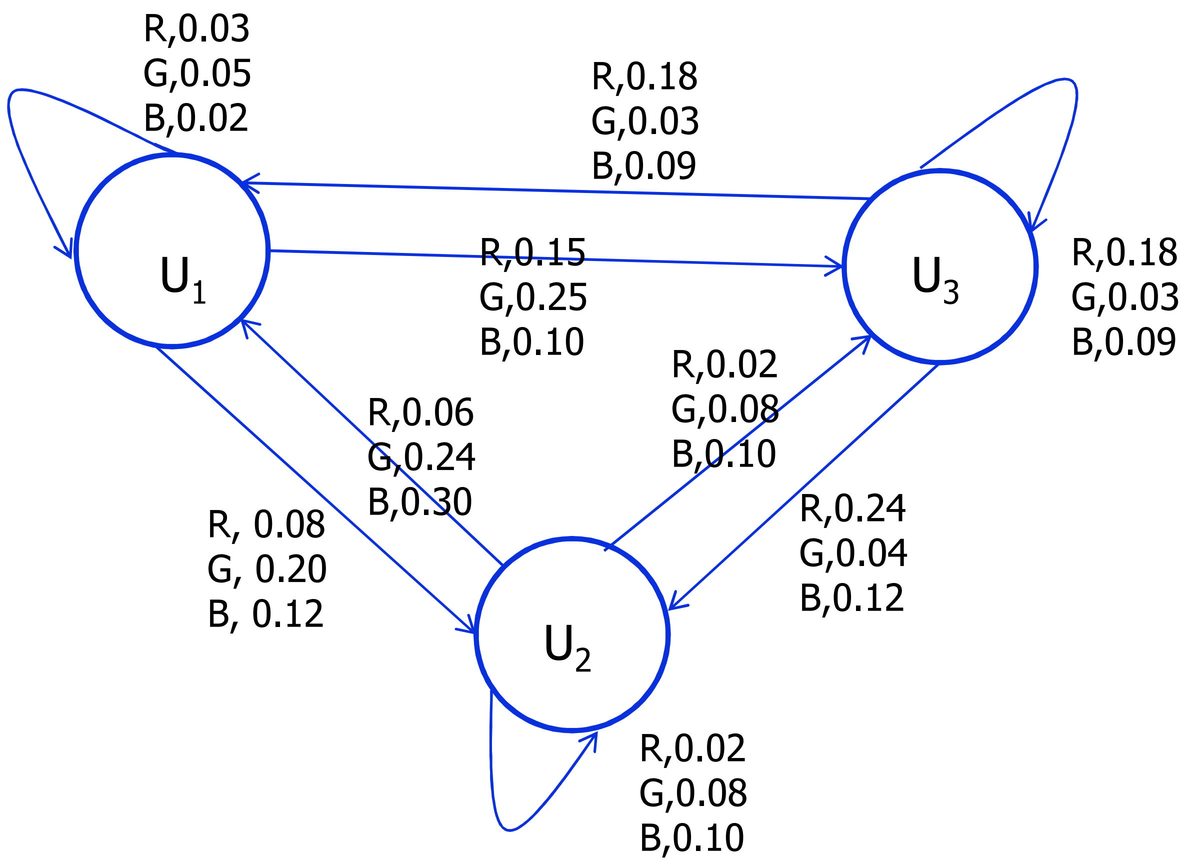
State Sequence : ??

Not so Easily Computable.

Diagrammatic representation (1/2)



Diagrammatic representation (2/2)



Example (contd.)

- Here :
  - S = {U1, U2, U3}
  - V = { R,G,B}
- For observation:
  - O ={o_1… o_n}
- And State sequence
  - Q ={q_1… q_n}
- π is $\pi_i = P(q_1 = U_i)$

A =

	U_1	U_2	U_3
U_1	0.1	0.4	0.5
U_2	0.6	0.2	0.2
U_3	0.3	0.4	0.3

B=

	R	G	B
U_1	0.3	0.5	0.2
U_2	0.1	0.4	0.5
U_3	0.6	0.1	0.3

Observations and states

	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
OBS:	R	R	G	G	B	R	G	R
State:	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8

$S_i = U_1/U_2/U_3$; A particular state

S: State sequence

O: Observation sequence

S* = “best” possible state (urn) sequence

Goal: Maximize P(S*|O) by choosing “best” S

Goal

- Maximize P(S|O) where S is the State Sequence and O is the Observation Sequence

$$S^* = \arg\max_S (P(S \mid O))$$

False Start

	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
OBS:	R	R	G	G	B	R	G	R
State:	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8

$$P(S|O) = P(S_{1-8}|O_{1-8})$$

$$P(S|O) = P(S_1|O).P(S_2|S_1,O).P(S_3|S_{1-2},O)\ldots P(S_8|S_{1-7},O)$$

By Markov Assumption (a state depends only on the previous state)

$$P(S|O) = P(S_1|O).P(S_2|S_1,O).P(S_3|S_2,O)\ldots P(S_8|S_7,O)$$

Baye's Theorem

$$P(A|B) = P(A).P(B|A)/P(B)$$

P(A) -: Prior
P(B|A) -: Likelihood

$$\text{argmax}_S P(S|O) = \text{argmax}_S P(S).P(O|S)$$

State Transitions Probability

$$P(S) = P(S_{1-8})$$

$$P(S) = P(S_1).P(S_2 | S_1).P(S_3 | S_{1-2}).P(S_4 | S_{1-3})..P(S_8 | S_{1-7})$$

By Markov Assumption (k=1)

$$P(S) = P(S_1).P(S_2 | S_1).P(S_3 | S_2).P(S_4 | S_3)..P(S_8 | S_7)$$

Observation Sequence probability

$$P(O|S)=P(O_1|S_{1-8}).P(O_2|O_1,S_{1-8}).P(O_3|O_{1-2},S_{1-8})..P(O_8|O_{1-7},S_{1-8})$$

Assumption that ball drawn depends only on the Urn chosen

$$P(O|S)=P(O_1|S_1).P(O_2|S_2).P(O_3|S_3)...P(O_8|S_8)$$

$$P(S|O)=P(S).P(O|S)$$

$$P(S|O)=P(S_1).P(S_2|S_1).P(S_3|S_2).P(S_4|S_3)...P(S_8|S_7).$$

$$P(O_1|S_1).P(O_2|S_2).P(O_3|S_3)...P(O_8|S_8)$$

Grouping terms

	O_0	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	
Obs:	ε	R	R	G	G	B	R	G	R	
State:	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9

$$
\begin{aligned}
&P(S).P(O|S) \\
&= [P(O_0|S_0).P(S_1|S_0)]. \\
&\quad [P(O_1|S_1).\quad P(S_2|S_1)]. \\
&\quad [P(O_2|S_2).\quad P(S_3|S_2)]. \\
&\quad [P(O_3|S_3).P(S_4|S_3)]. \\
&\quad [P(O_4|S_4).P(S_5|S_4)]. \\
&\quad [P(O_5|S_5).P(S_6|S_5)]. \\
&\quad [P(O_6|S_6).P(S_7|S_6)]. \\
&\quad [P(O_7|S_7).P(S_8|S_7)]. \\
&\quad [P(O_8|S_8).P(S_9|S_8)].
\end{aligned}
$$

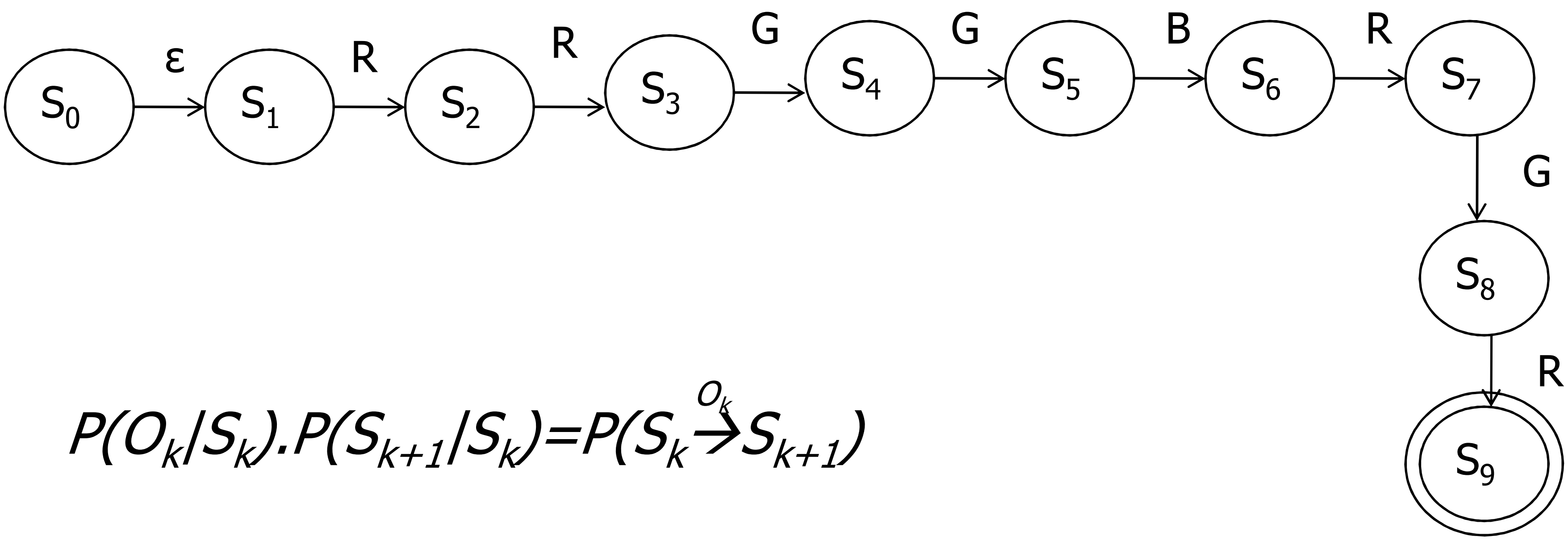
We introduce the states S_0 and S_9 as initial and final states respectively.

After S_8 the next state is S_9 with probability 1, i.e., $P(S_9|S_8)=1$

O_0 is ε-transition

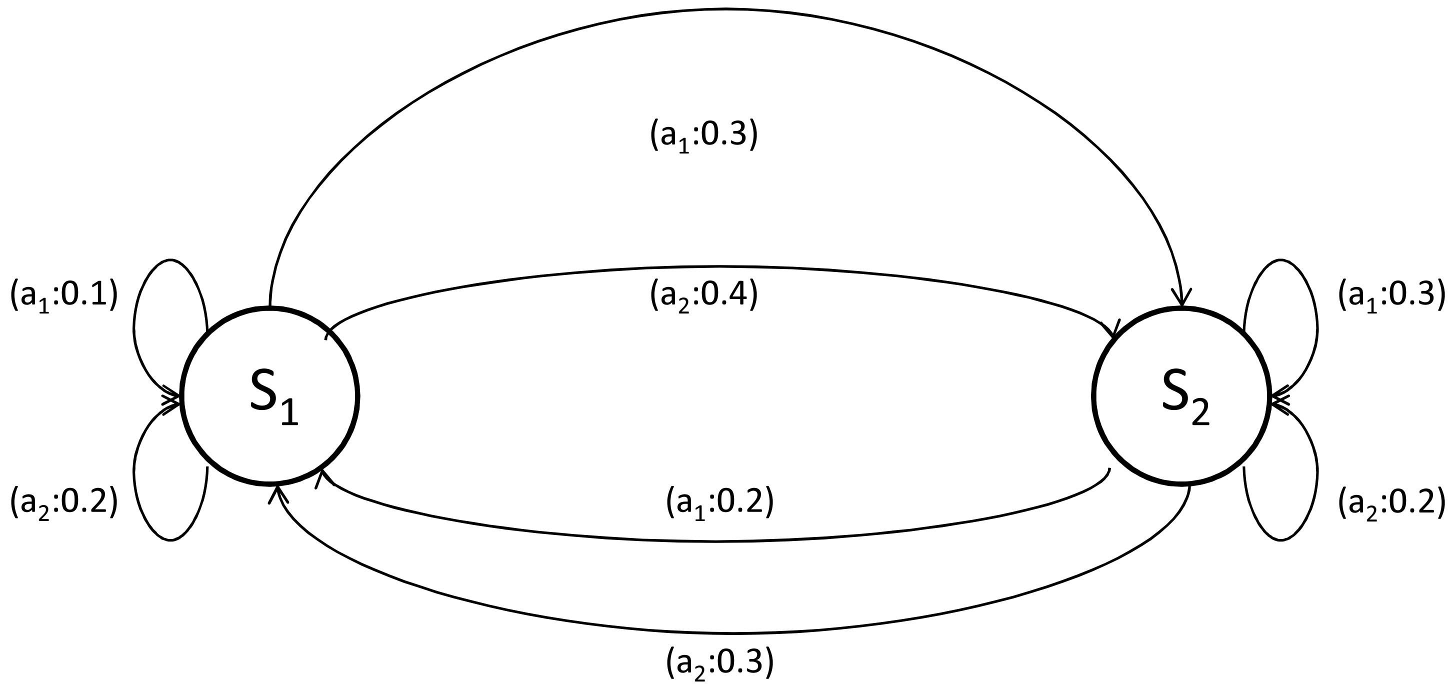
Introducing useful notation

	O_0	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	
Obs:	ε	R	R	G	G	B	R	G	R	
State:	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9



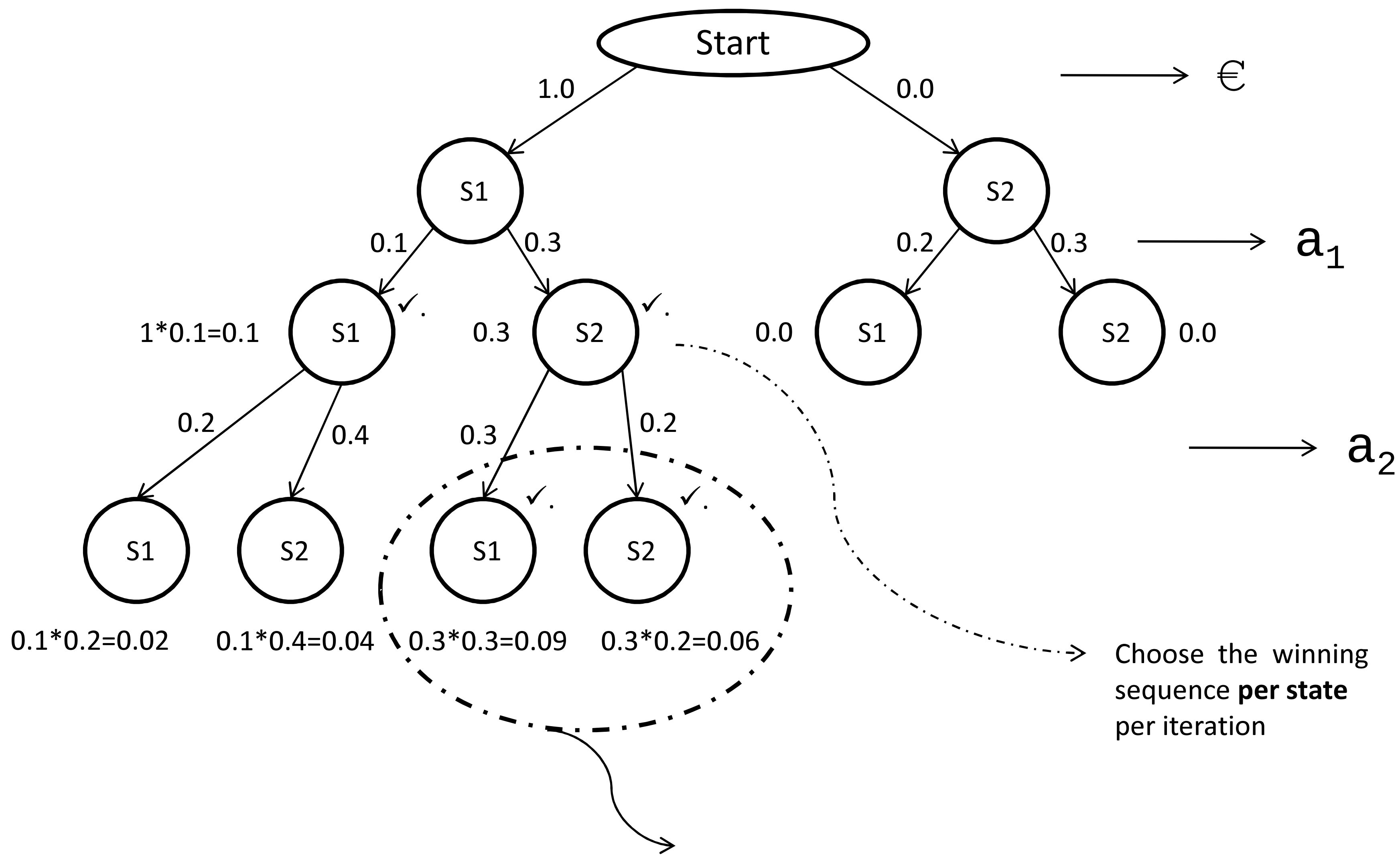
$$P(O_k|S_k).P(S_{k+1}|S_k)=P(S_k \xrightarrow{O_k} S_{k+1})$$

Probabilistic FSM

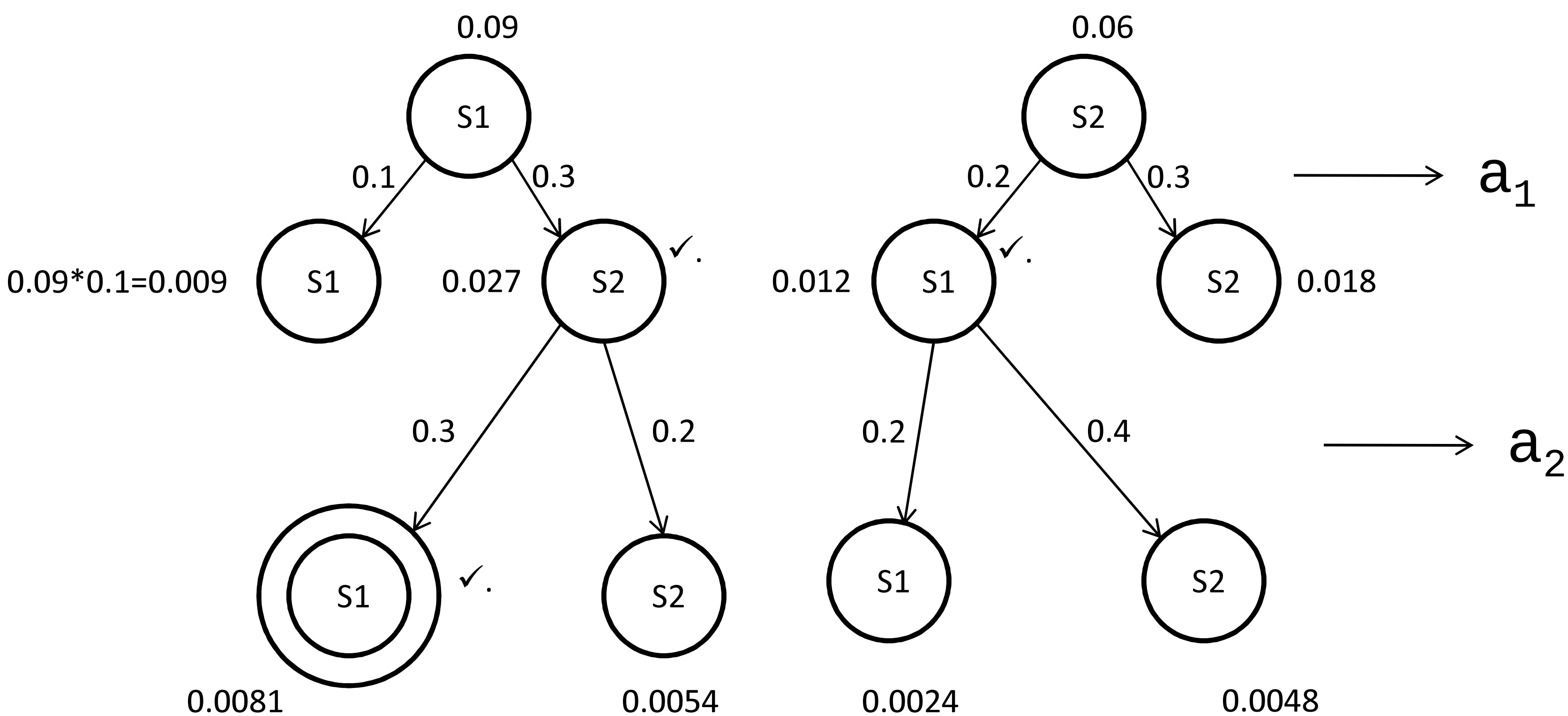


The question here is:
"what is the most likely state sequence given the output sequence seen"

Developing the tree



Tree structure contd…



The problem being addressed by this tree is $S^* = \arg\max_{S} P(S \mid a_1 - a_2 - a_1 - a_2, \mu)$

a1-a2-a1-a2 is the output sequence and μ the model or the machine

Path found:
(working backward)

$S_1 \xrightarrow{a_1} S_2 \xrightarrow{a_2} S_1 \xrightarrow{a_1} S_2 \xrightarrow{a_2} S_1$

Problem statement: Find the best possible sequence

$$S^* = \arg\max_{S} P(S \mid O, \mu)$$

$$where,\ S \rightarrow \text{State Seq},\ O \rightarrow \text{Output Seq},\ \mu \rightarrow \text{Model or Machine}$$

$$\text{Model or Machine} = \{S_0, S, A, T\}$$

- S_0 → Start symbol
- S → State collection
- A → Alphabet set
- T → Transitions

T is defined as $P(S_i \xrightarrow{a_k} S_j) \quad \forall_{i,\ j,\ k}$