Artificial & Computational Intelligence

AIML CLZG557

M6: Reasoning over time

BITS Pilani
Pilani Campus

Dr. Sudheer Reddy

Course Plan

M1	Introduction to AI
M2	Problem Solving Agent using Search
M3	Game Playing
M4	Knowledge Representation using Logics
M5	Probabilistic Representation and Reasoning
M6	Reasoning over time
M7	Ethics in AI

Module 6:
Reasoning over time

Reasoning Over Time

A. Time and Uncertainty

B. Inference in temporal models

C. Overview of HMM

D. Learning HMM Parameters using EM Algorithm

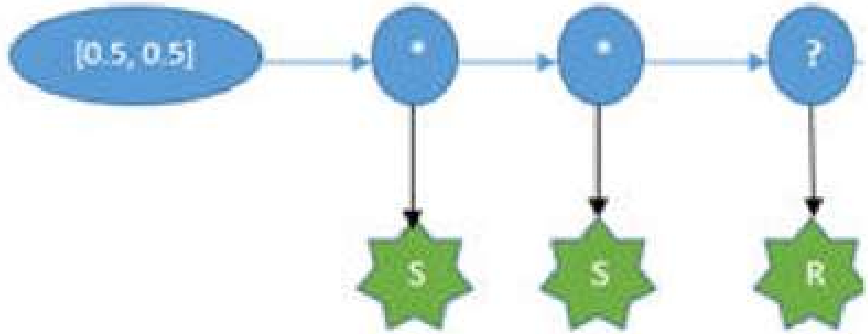
E. Applications of HMM

Hidden Morkov Model

Filtering : Forward Propagation Algorithm

Find the Current Pressure if sequence of weather observations recorded are: **S-S-R**

Intuition: $P(E_{1...t}) = \sum_{i=1}^{N} P(E_{1...t} \mid X_{1...t}) * P(X_{1...t}) = \sum_{i=1}^{N} \prod_{j=1}^{t} P(Ej \mid Xj) * P(Xj \mid Xj_{-1})$



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

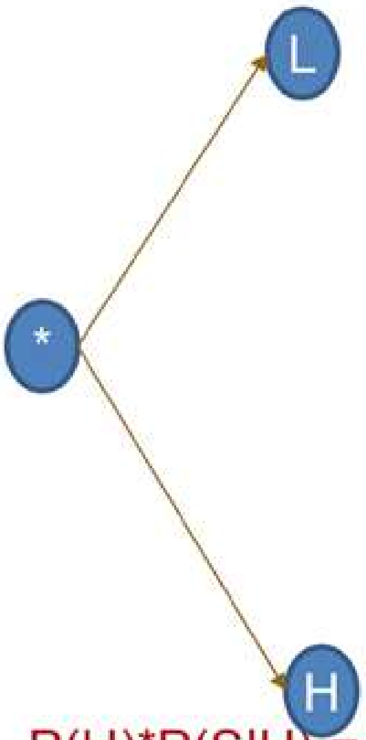
Hidden Morkov Model

Forward Propagation Algorithm

Pressure sequence observation: **S-S-R**

Initialization Phase:

$P(L)*P(S|L) = 0.5*0.2 = 0.1 \rightarrow 0.25$



$P(H)*P(S|H) = 0.5*0.6 = 0.3 \rightarrow 0.75$

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

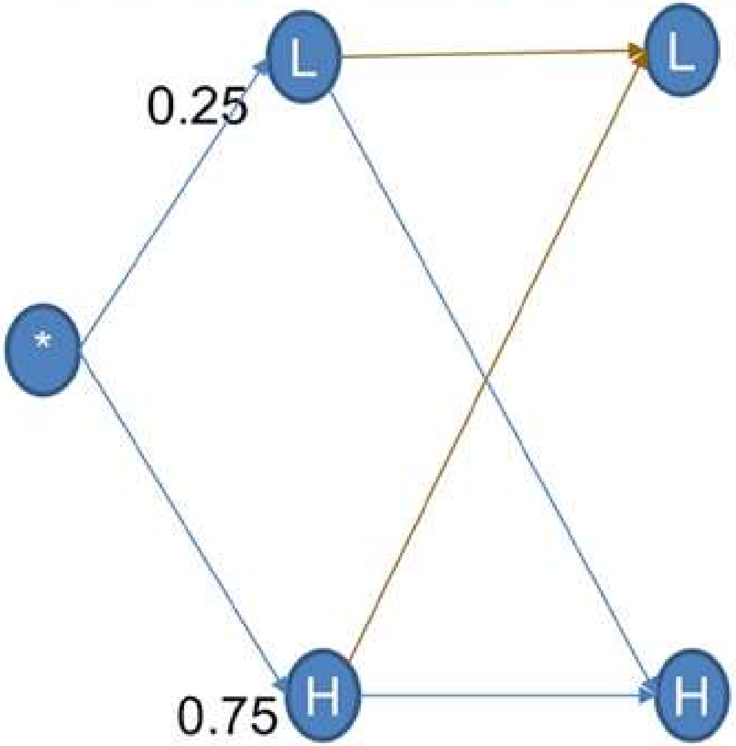
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

P(L)*P(L|L)*P(S|L) = 0.25*0.5*0.2 = **0.025**

P(H)*P(L|H)*P(S|L) = 0.75*0.2*0.2 = **0.03**



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

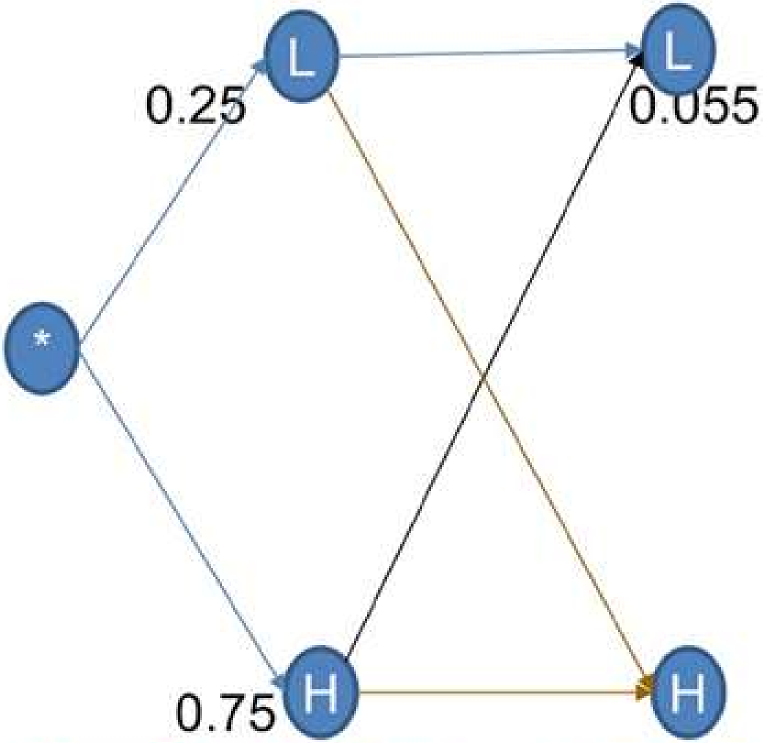
Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

<u>Recursion Phase:</u>

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R



P(L)*P(H|L)*P(S|H) = 0.25*0.5*0.6 = **0.075**

P(H)*P(H|H)*P(S|H) = 0.75*0.8*0.6 = **0.36**

Transition Model / Probability Matrix

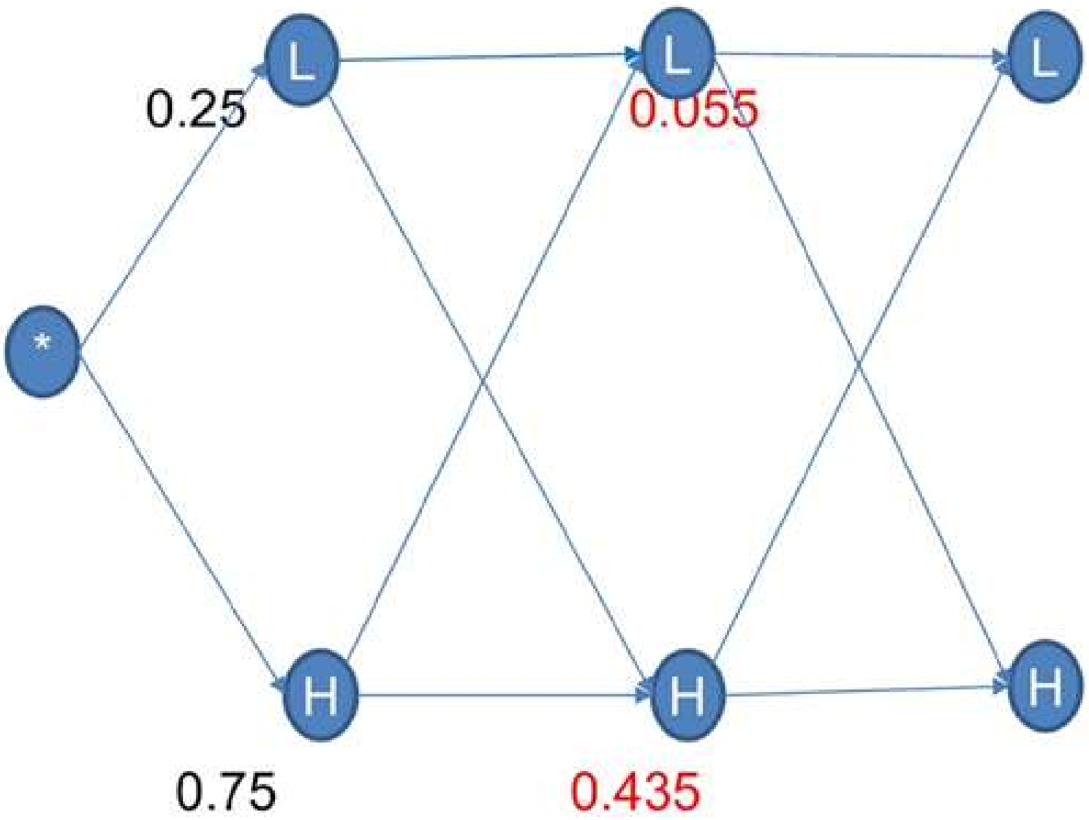
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

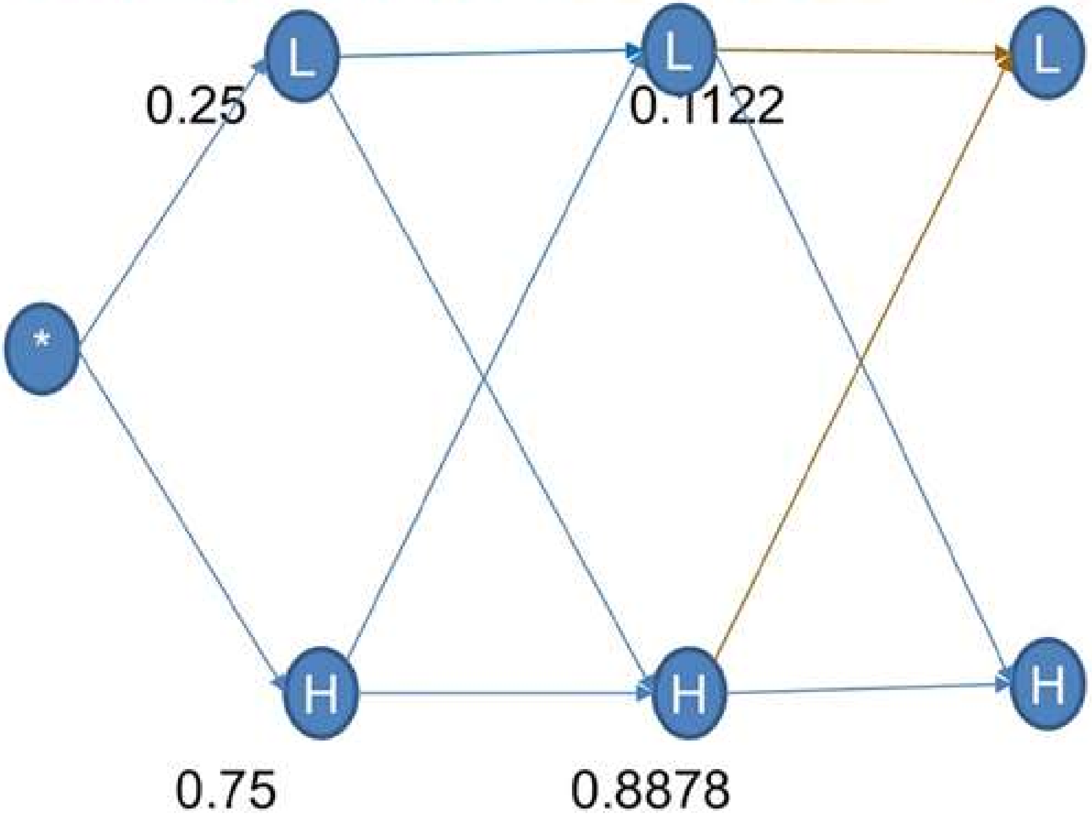
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

P(L)*P(L|L)*P(R|L) = 0.1122*0.5*0.8 = **0.04488**

P(H)*P(L|H)*P(R|L) = 0.8878*0.2*0.8 = **0.142048**



Transition Model / Probability Matrix

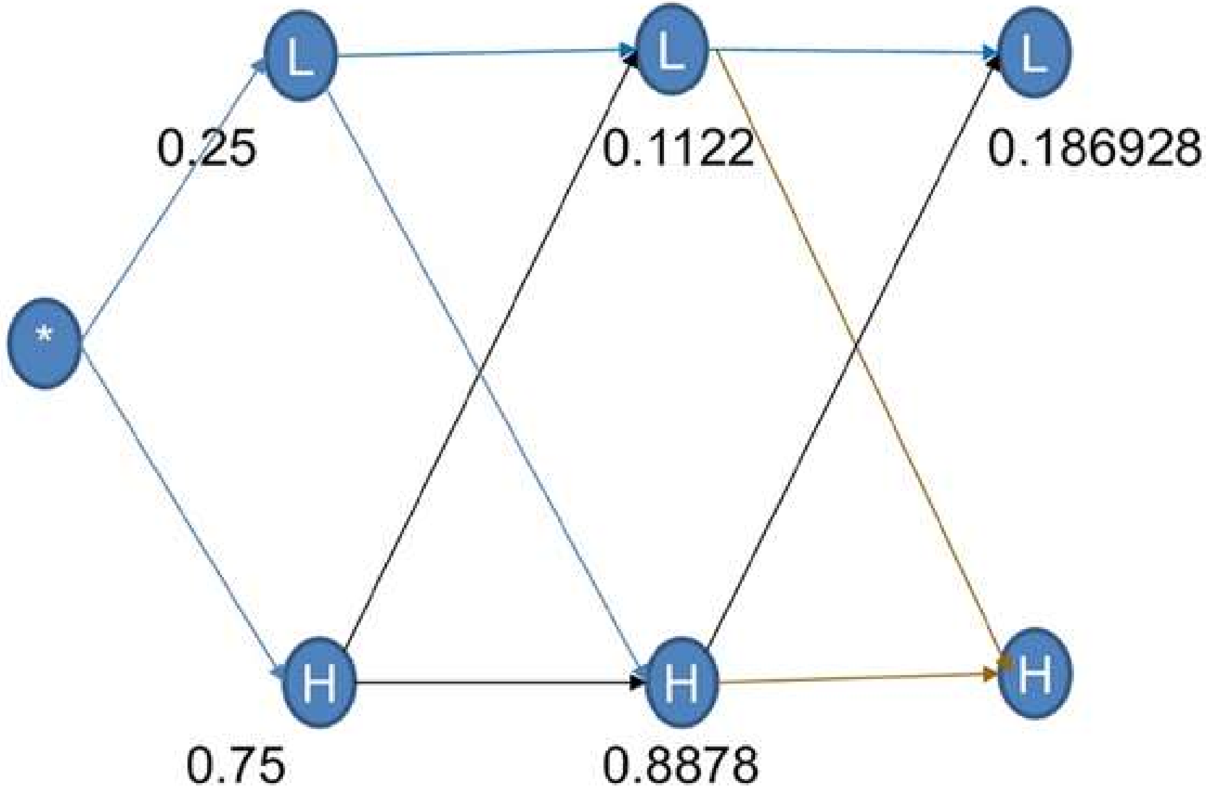
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R



P(L)*P(H|L)*P(R|H) = 0.1122*0.5*0.4 = **0.02244**

P(H)*P(H|H)*P(R|H) = 0.8878*0.8*0.4 = **0.284096**

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

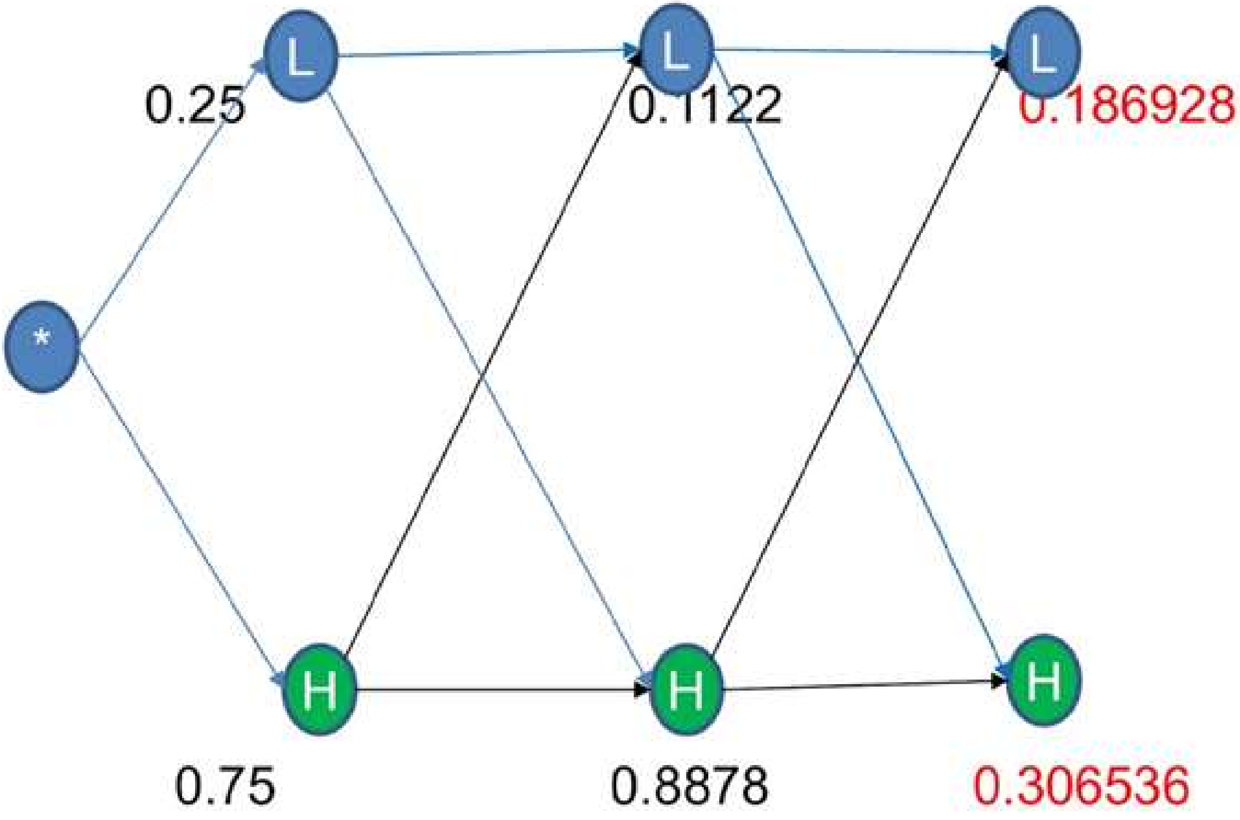
Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

Termination Phase:



Transition Model / Probability Matrix

P(U_{t-1} = HP)	P(U_{t-1} = LP)	←Previous
0.2	0.5	P(U_t = LP)
0.8	0.5	P(U_t = HP)

Evidence / Sensor Model/ Emission Probability Matrix

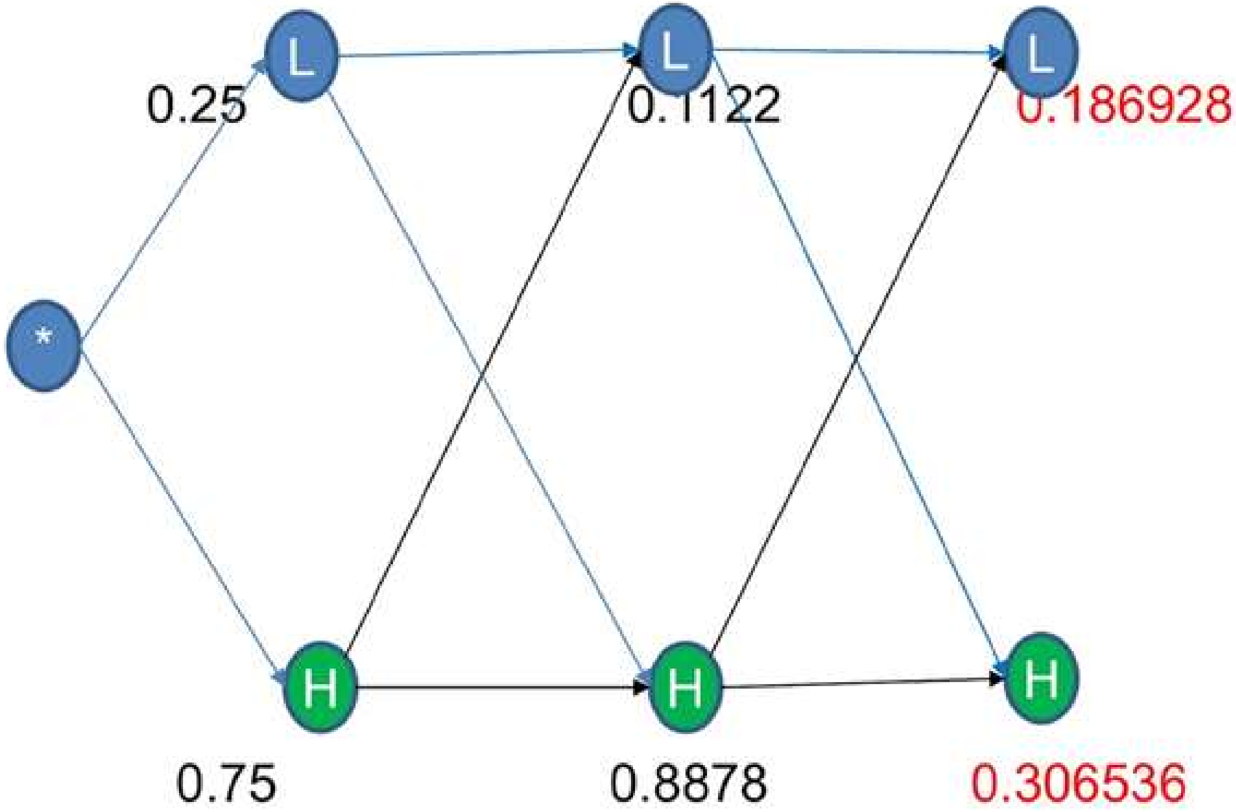
P(X_t = LP)	P(X_t = HP)	←Unobserved Evidence v
0.8	0.4	P(E_t = Rainy)
0.2	0.6	P(E_t = Sunny)

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

<u>Termination Phase:</u>

(0.37881,**0.62119**)



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

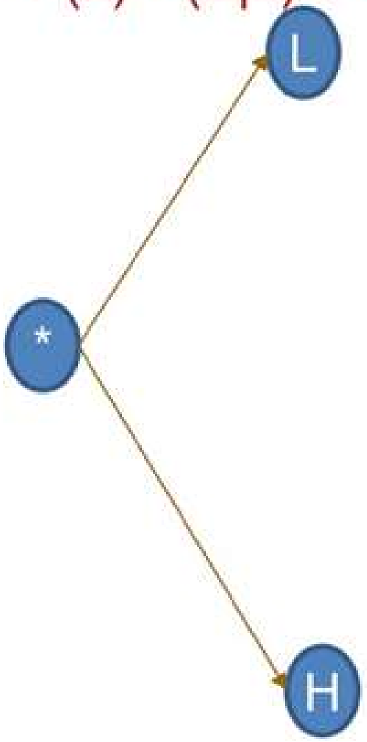
Most Likely Explanation : Veterbi Algorithm

Find the pattern in pressure that might have caused this observation: **S-S-R**

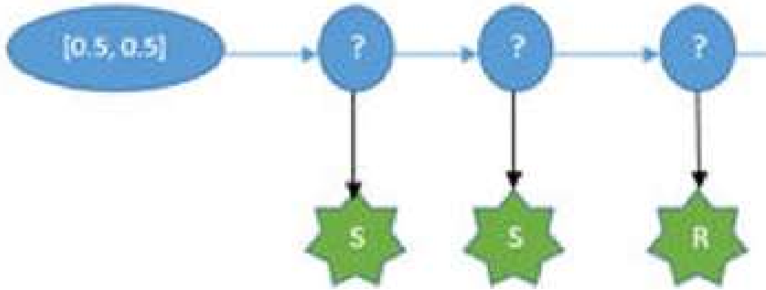
argmax $X_{1...t}$: $P(X_{1...t} \mid E_{1...t})$

MM Inf

P(L)*P(S|L) = 0.5*0.2 = 0.1 → 0.25



P(H)*P(S|H) = 0.5*0.6 = 0.3 → 0.75



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

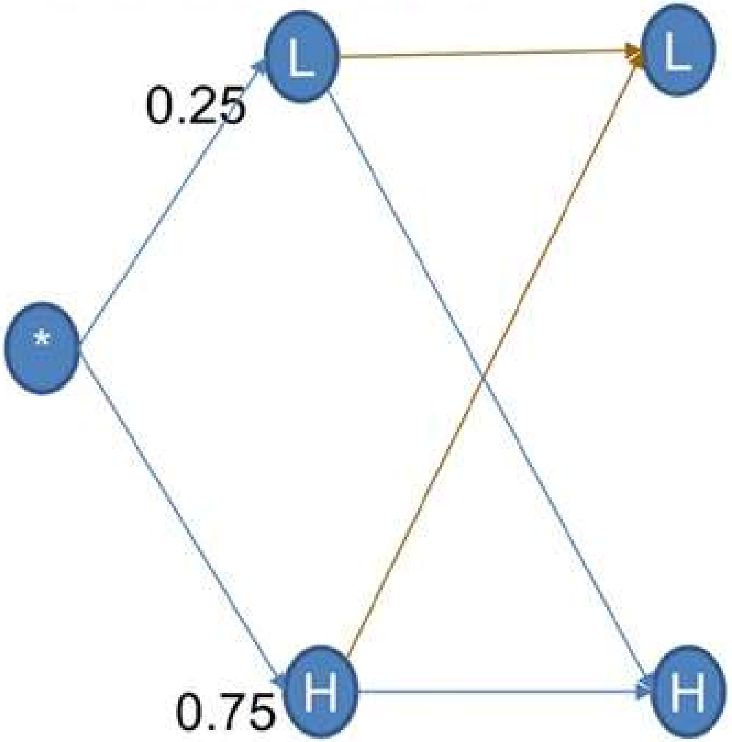
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Veterbi Algorithm : S-S-R

P(L)*P(L|L)*P(S|L) = 0.25*0.5*0.2 = 0.025

P(H)*P(L|H)*P(S|L) = 0.75*0.2*0.2 = **0.03**



Transition Model / Probability Matrix

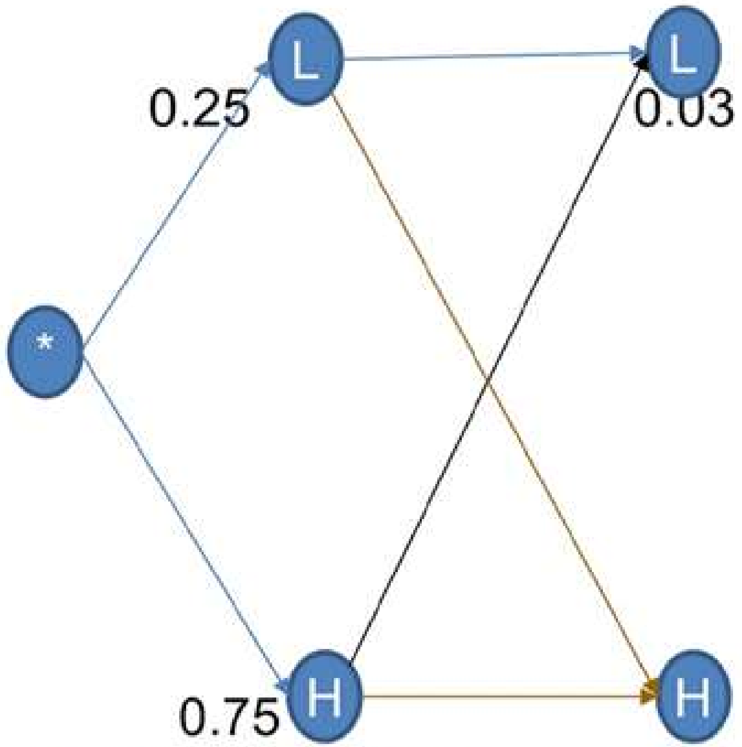
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Veterbi Algorithm : S-S-R



P(L)*P(H|L)*P(S|H) = 0.25*0.5*0.6 = 0.075

P(H)*P(H|H)*P(S|H) = 0.75*0.8*0.6 = **0.36**

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	← Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

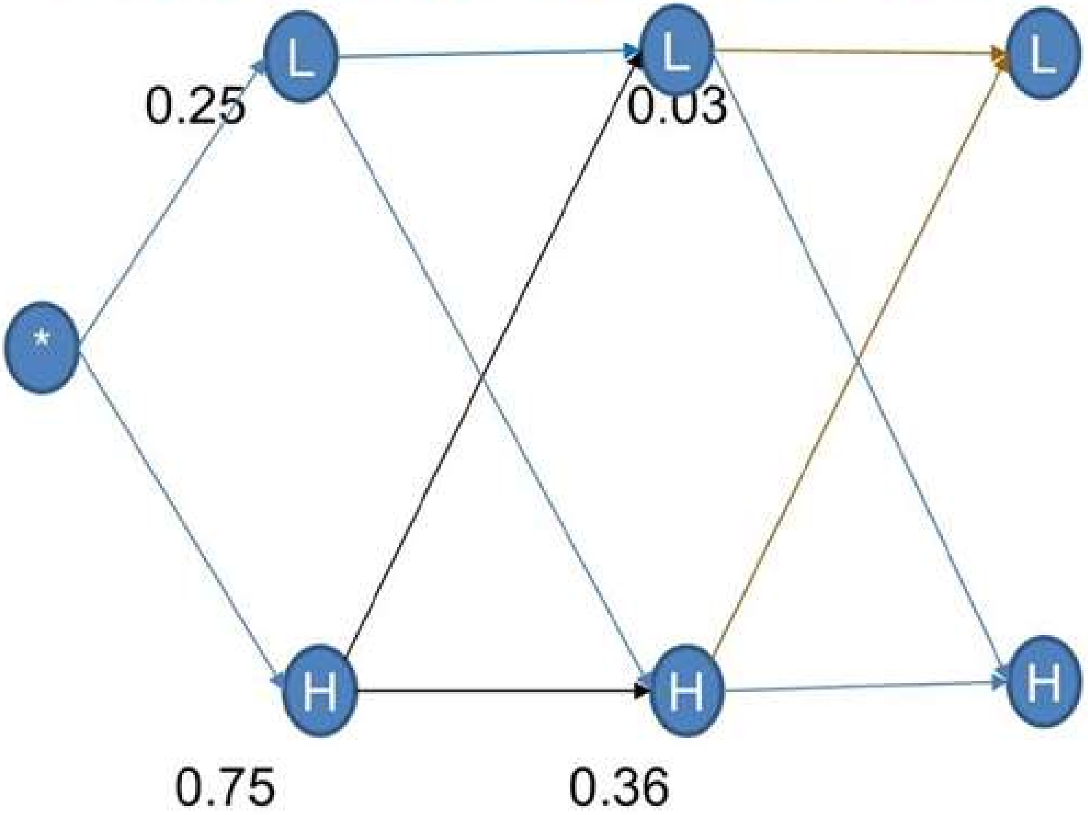
$P(X_t = LP)$	$P(X_t = HP)$	← Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Veterbi Algorithm : S-S-R

P(L)*P(L|L)*P(R|L) = 0.03*0.5*0.8 = 0.012

P(H)*P(L|H)*P(R|L) = 0.36*0.2*0.8 = **0.0576**



Transition Model / Probability Matrix

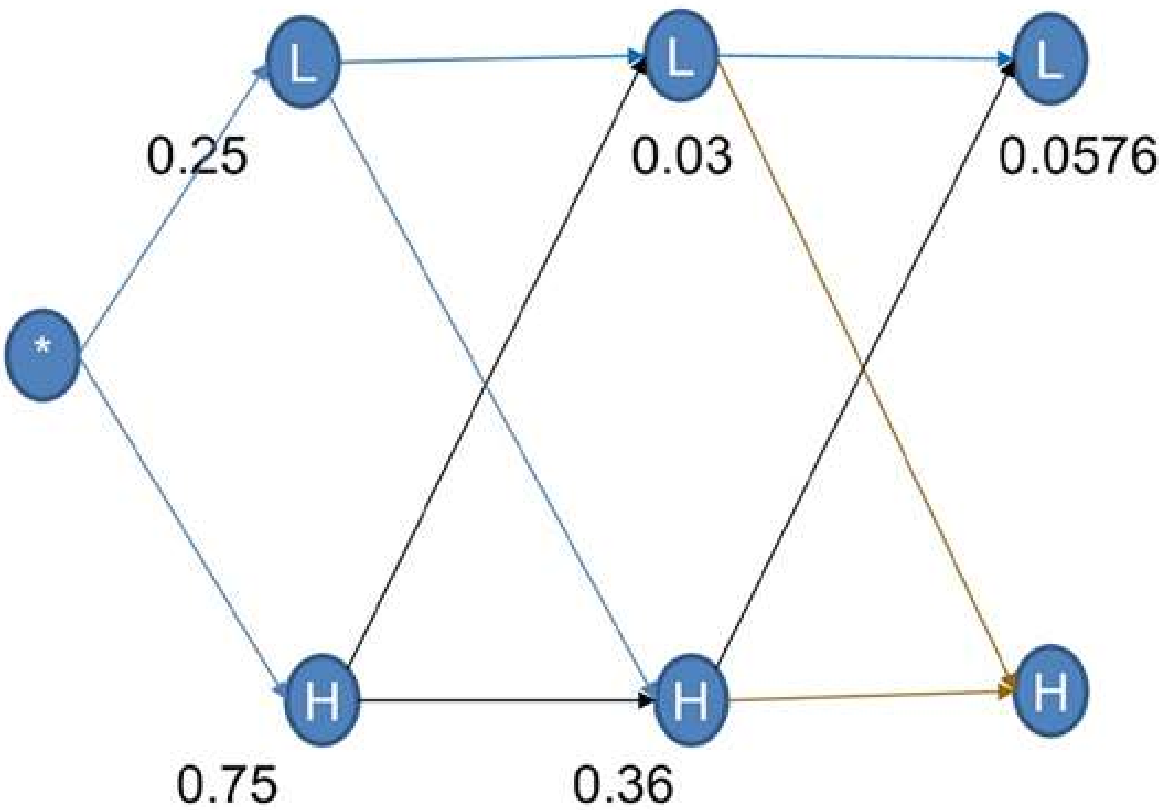
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Veterbi Algorithm : S-S-R



P(L)*P(H|L)*P(R|H) = 0.03*0.5*0.4 = 0.006

P(H)*P(H|H)*P(R|H) = 0.36*0.8*0.4 = **0.1152**

Transition Model / Probability Matrix

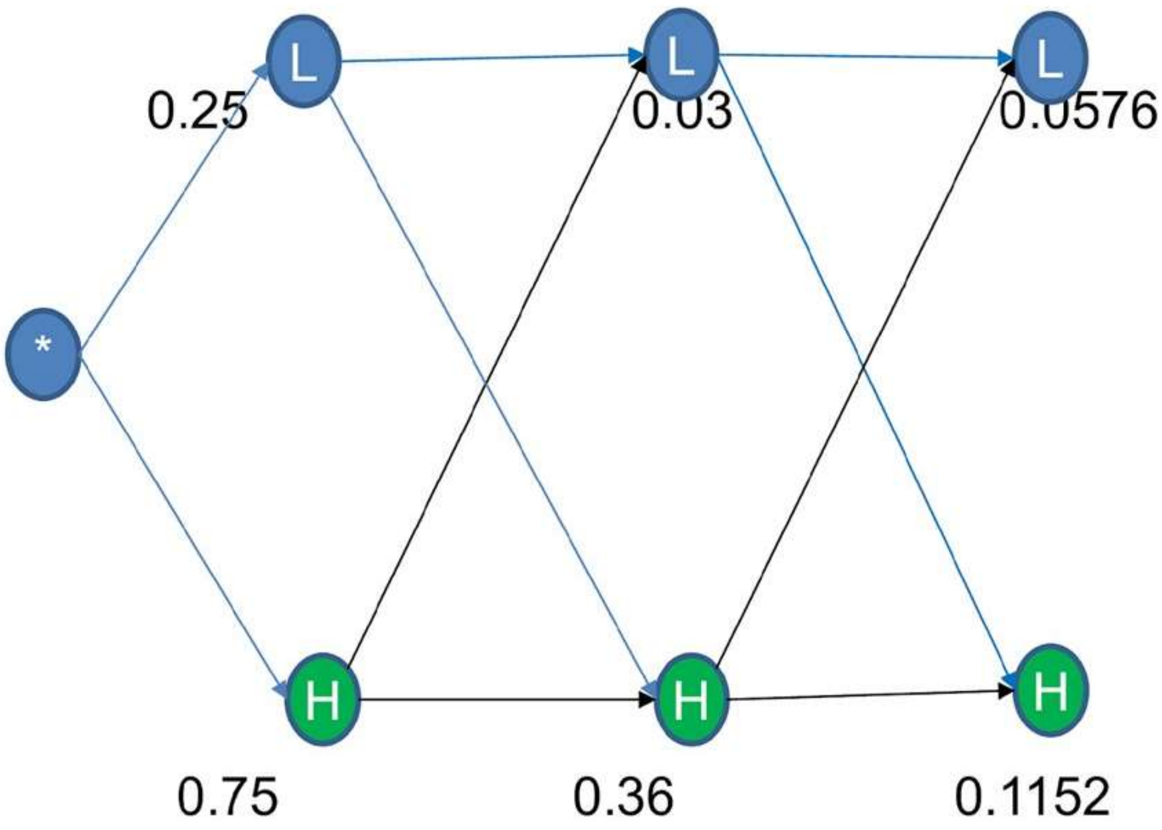
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

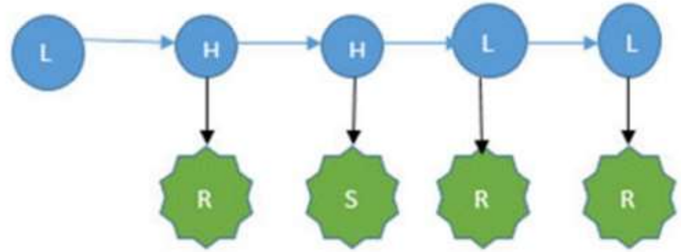
Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Veterbi Algorithm : S-S-R





Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

innovate achieve lead

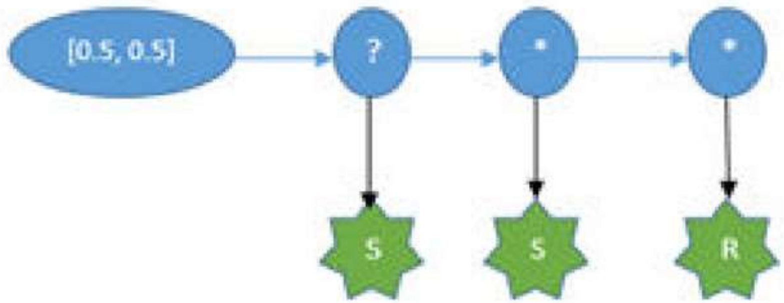
Hidden Morkov Model

Inference: Type -4

Smoothing : Backward Propagation Algorithm (Most Likely State Estimation)

Find the Pressure in past instance of time if sequence of following future weather observations recorded are: **S-S-R**

Intuition: $P(E_{1\ldots t}) = \sum_{i=1}^{N} P(E_{1\ldots t} \mid X_{1\ldots t}) * P(X_{1\ldots t}) = \sum_{i=1}^{N} \prod_{j=1}^{t} P(Ej \mid Xj) * P(Xj \mid Xj_{-1})$



Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

Inference: Type -4

Smoothing : Backward Propagation Algorithm

Find the Pressure in past instance of time if sequence of following future weather observations recorded are: **S-S-R**

Intuition: $P(Xt_{+1} | E_{1...t+1}) = \boldsymbol{\alpha}\, P(et_{+1} | Xt_{+1}) * \sum_{X_t} P(X_{t+1}|X_t) * \boldsymbol{P(X_t | E_{1..t})}$

$P(X_1 | SSR) = P(X_1 | S,S,R)$

$$= \frac{\boldsymbol{P(SR\,|X_1\,S)} * P(X_1|\,S)}{P(SR)}$$

$$= \frac{P(X_1|\,S) * \{\sum_{X_2} P(X_2|X_1) * \boldsymbol{P(SR\,|\,X_2\,X_1)}\}}{P(SR)}$$

$$= \frac{P(X_1|\,S) * \{\sum_{X_2} P(X_2|X_1) * \boldsymbol{P(SR\,|\,X_2)}\}}{P(SR)}$$

$$= \frac{P(X_1|\,S) * \{\sum_{X_2} P(X_2|X_1) * P(S\,|\,X_2) * \boldsymbol{P(R\,|\,X_2)}\}}{P(SR)}$$

$$= \frac{P(X_1|\,S) * \{\sum_{X_2} P(X_2|X_1) * P(S\,|\,X_2) * \{\sum_{X_3} P(X_3|X_2) * P(R\,|\,X_3) * \boldsymbol{P(.\,|\,X_3)}\}\}}{P(SR)}$$

Transition Model / Probability Matrix

P(U$_{t-1}$ = HP)	P(U$_{t-1}$ = LP)	←Previous
0.2	0.5	P(U$_t$ = LP)
0.8	0.5	P(U$_t$ = HP)

Evidence / Sensor Model/ Emission Probab

P(X$_t$ = LP)	P(X$_t$ = HP)	←Unobserved Evidence v
0.8	0.4	P(E$_t$ = Rainy)
0.2	0.6	P(E$_t$ = Sunny)

$$P(Xt\,|E_{t+1,\,t+2\,...\,z}) = \boldsymbol{\alpha} * \text{fwd msg} * \sum_{X_{t+1}} P(X_{t+1}|X_t) * P(e_{t+1}|X_{t+1}) * \boldsymbol{P(E_{t+2..z}|\,X_{t+1})}$$

Hidden Morkov Model

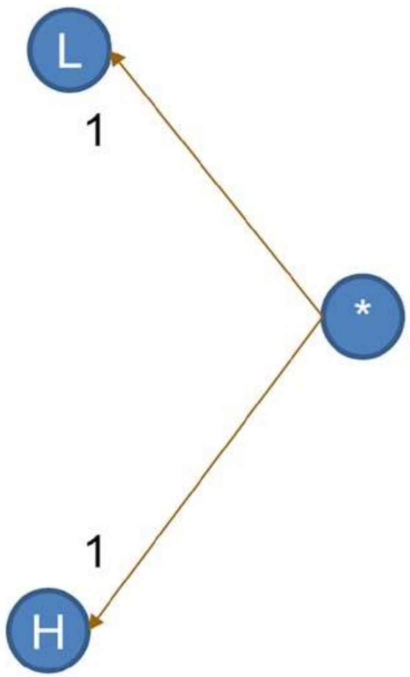
Backward Propagation Algorithm

Pressure sequence observation: **S-S-R**

Initialization Phase: Set value 1 for the terminal state

P(L|L)*P(R|L)*P(.|L) = 0.5*0.8 * 1= 0.40

P(H|L)*P(R|H)*P(.|H) = 0.5*0.4 *1= 0.2



P(L|H)*P(R|L)*P(.|L) = 0.2*0.8 * 1= 0.16

P(H|H)*P(R|H)*P(.|H) = 0.8*0.4 *1= 0.32

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

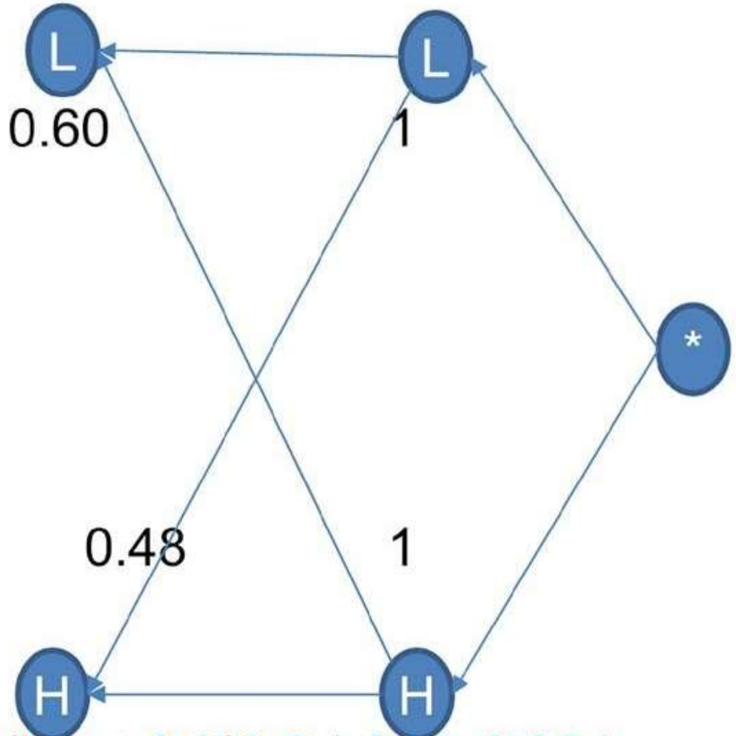
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Backward Propagation Algorithm : S-S-R

P(L|L)*P(S|L)*MSG(L`) = 0.5*0.2 * 0.60= 0.06

P(H|L)*P(S|H)*MSG(H`) = 0.5*0.6*0.48= 0.144



P(L|H)*P(S|L)*MSG(L`) = 0.2*0.2 * 0.6= 0.024

P(H|H)*P(S|H)*MSG(H`) = 0.8*0.6 *0.48= 0.2304

Recursion Phase:

Transition Model / Probability Matrix

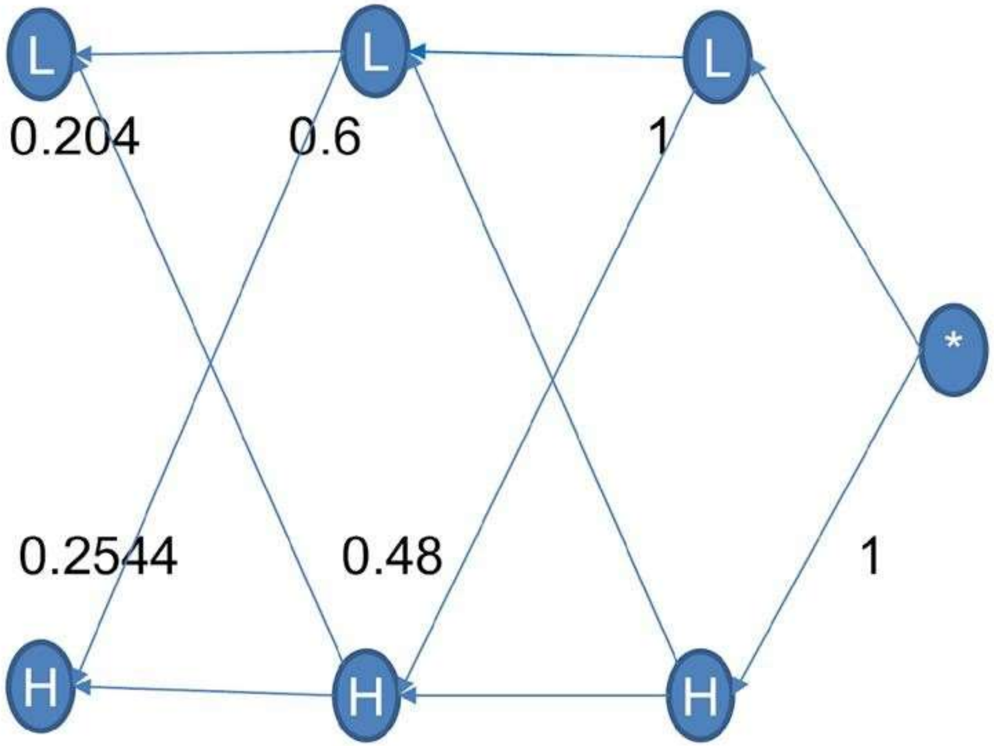
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Backward Propagation Algorithm : S-S-R



Recursion Phase: If it continues if needed !!!!

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

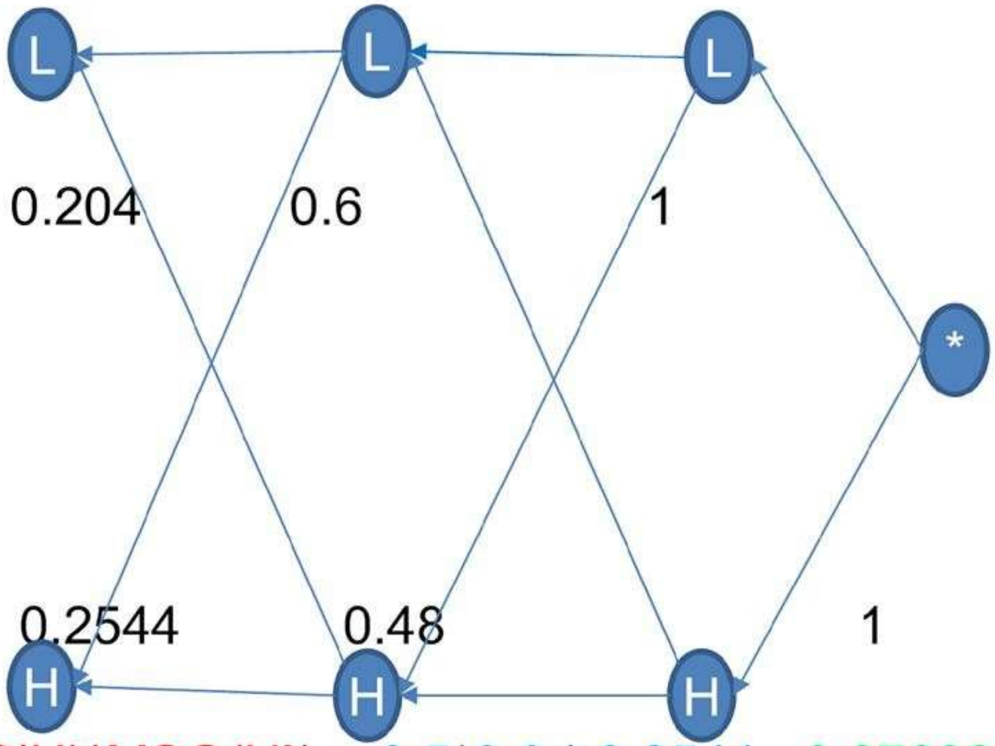
Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Backward Propagation Algorithm : S-S-R

P(L)*P(S|L)*MSG(L`) = 0.5*0.2 * 0.204= 0.0204



P(H)*P(S|H)*MSG(H`) = 0.5*0.6 * 0.2544= 0.07632

<u>Termination Phase: </u>(0.2109,0.7891)

Normalize :Initial value * Emission at start* backMsg

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

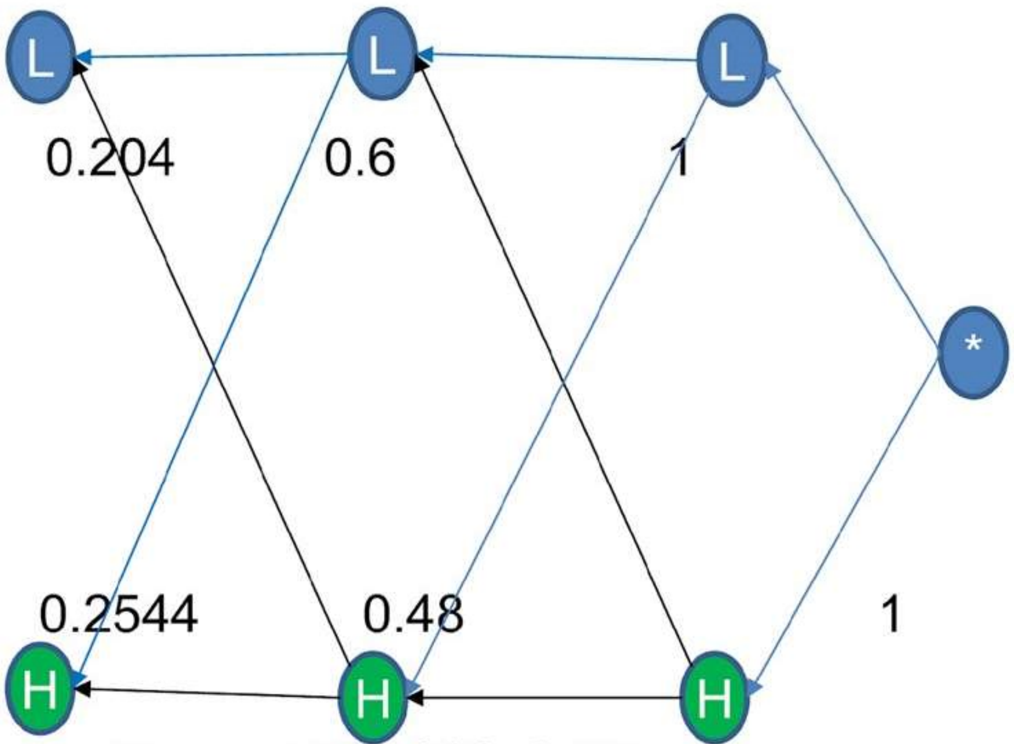
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Forward Backward Propagation Algorithm : S-S-R

$P(X2 \mid SSR) = \alpha * P(X2|SS) * P(R|X2)$

$P(X2 \mid SSR) = \alpha * (0.1122 , 0.8878) * (0.6, 0.48) = (0.06732, 0.426144) = (0.14,0.86)$



Termination Phase: X2 = ??? → **X2 = H**

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Forward Path Probability

$\alpha_t(j) = \alpha_{t-1}(i)\, a_{i,j}\, b_j(o_t)$ with summation index i printed below

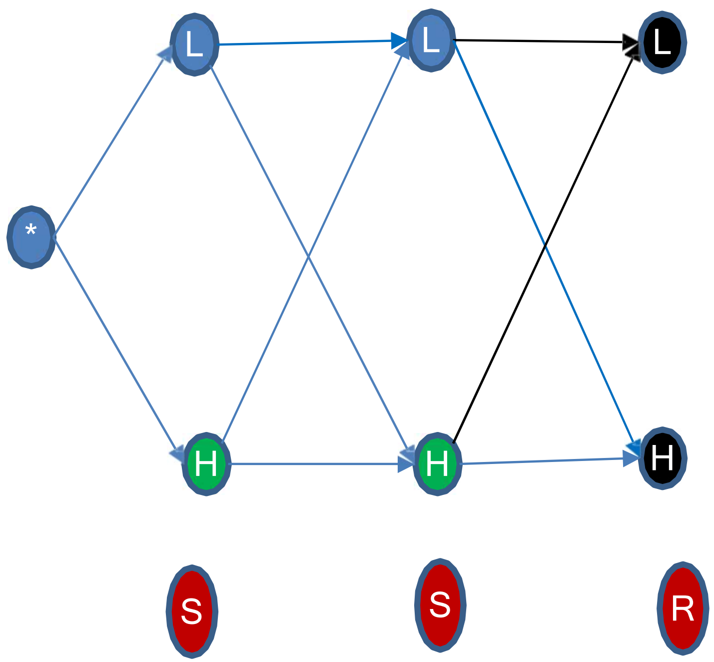
$P(O_{1..t} \mid \lambda)$

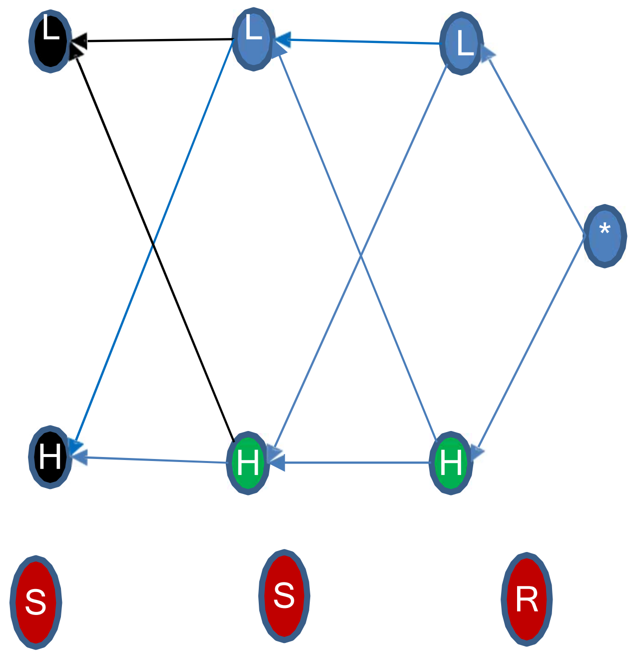
Backward Path Probability

$\beta_t(i) = \beta_{t+1}(j)\, a_{i,j}\, b_j(o_{t+1})$ with summation index j printed below

$P(O_{t+1,\, t+2..t+k} \mid \lambda)$

$\gamma_t(i) = P(X_t \mid O_{1...t\ ...t+1,\ t+2..t+k} \mid \lambda)$: Forward - Backward Algorithm





```
function FORWARD-BACKWARD(ev, prior) returns a vector of probability distributions
  inputs: ev, a vector of evidence values for steps 1, ..., t
          prior, the prior distribution on the initial state, P(X_0)
  local variables: fv, a vector of forward messages for steps 0, ..., t
                   b, a representation of the backward message, initially all 1s
                   sv, a vector of smoothed estimates for steps 1, ..., t

  fv[0] ← prior
  for i = 1 to t do
      fv[i] ← FORWARD(fv[i − 1], ev[i])
  for i = t downto 1 do
      sv[i] ← NORMALIZE(fv[i] × b)
      b ← BACKWARD(b, ev[i])
  return sv
```

Figure 15.4 The forward–backward algorithm for smoothing: computing posterior probabilities of a sequence of states given a sequence of observations. The FORWARD and BACKWARD operators are defined by Equations (15.5) and (15.9), respectively.

Text & Natural Language Processing

HMM Application

Initial	Prob
N	0.67
D	0.33
V	0
J	0
A	0
P	0

N	D	V	J	A	P	
	01675		0.67		1	N
		0.571				D
0.63	0.1675	0.143				V
	0.33	0.143				J
						A
	0.1675					P
0.37	0.1675	0.143	0.33			E

Given the corpus with tags to build training data:

1. Create initial probability matrix.
2. Transition probability matrix
3. Emission probability matrix
4. Use HMM Veterbi algorithm to predict the sequence of PoS Tags for given test data / sentence.

In the HMM model , the PoS tags act as the hidden states and the word in the given test sentence as the observed states.

- Boys are taller.
  N V J
- This is the tree.
  D V D N
- She is a tall girl.
  N V D J N
- Trees are more.
  N V D
- Girls are more than boys.
  N V D P N
- The tall tree is falling.
  D J N V V

Initi Prob al	
N	0.67
D	0.33
V	0
J	0
A	0
P	0

N	D	V	J	A	P	
	0.1675		0.67		1	N
		0.571				
0.63	0.1675	0.143				
	0.33	0.143				
	0.1675					
0.37	0.1675	0.143				
		0.33				

N	D	V	J	A	P	
0.25						Boys
		0.43				Are
			1			Tall
	0.17					This
		0.43				Is
	0.33					The
0.375						Tree
0.125						She
	0.17					A
0.25						Girl
	0.33					More
					1	Than
		0.14				fall

- Boys are taller.
  N V J
- This is the tree.
  D V D N
- She is a tall girl.
  N V D J N
- Trees are more.
  N V D
- Girls are more than boys.
  N V D P N
- The tall tree is falling.
  D J N V V

Initi Prob	al
N	0.67
D	0.33
V	0
J	0
A	0
P	0

N	D	V	J	A	P	
	01675		0.67		1	N
		0.571				
0.63	0.1675	0.143				
	0.33	0.143				
	0.1675					
0.37	0.1675	0.143				
		0.33				

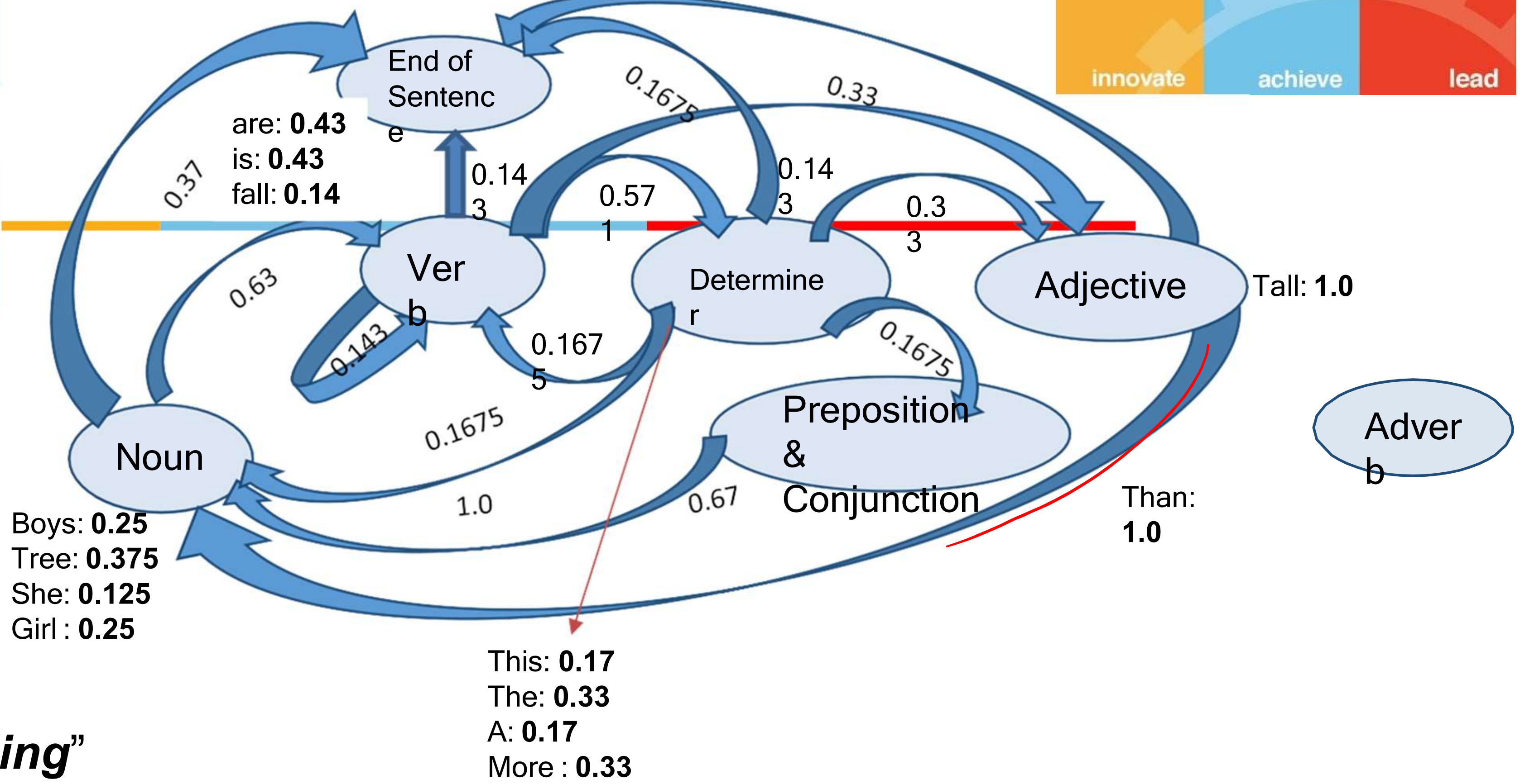
N	D	V	J	A	P	
0.25						Boys
		0.43				Are
			1			Tall
	0.17					This
		0.43				Is
	0.33					The
0.375						Tree
0.125						She
	0.17					A
0.25						Girl
	0.33					More
					1	Than
		0.14				fall

Exercise :

For the below test data/sentence, using the tables constructed using training data, predict the PoS tags.

"Girls are falling"

Initi	
N	0.67
D	0.33
V	0
J	0
A	0
P	0



"Girls are falling"

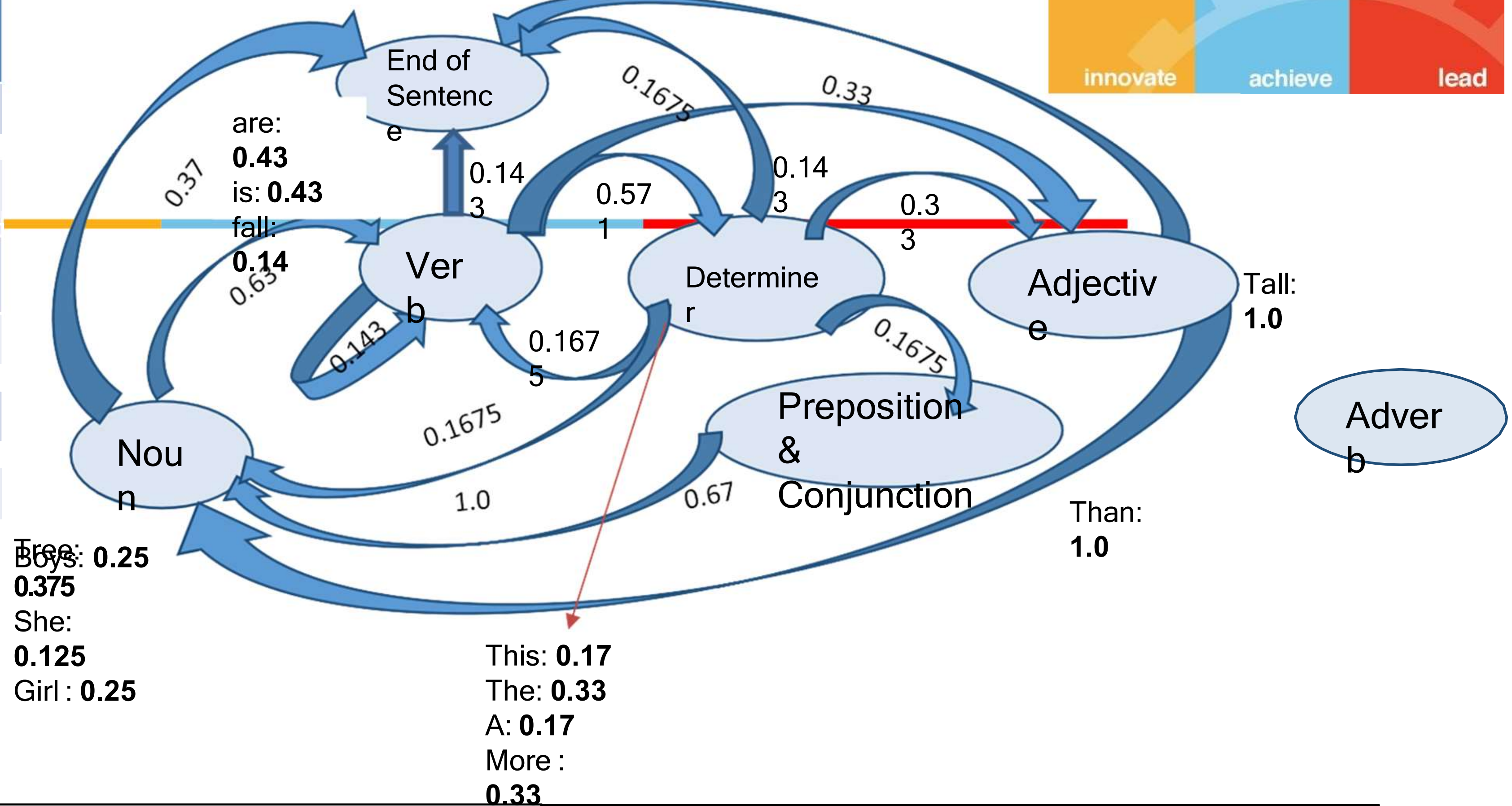
P(Girls , Noun) = P(Girl | Noun) * P(Noun | StartState) = 0.25 * 0.67 = **0.1675**

P(Girls , Verb) = P(Girl | Verb) * P(Verb | StartState) = 0 * 0 = 0 (Ideally with better corpus and the KB , for most cases it might not be 0 but too low like 0.00000000000…..001.)

P(Girls , Determiner) = P(Girls , Adverb) = P(Girls , Adjective) =P(Girls , Preposition/Conjunction) = 0

StartState → Noun

Initi	
N	0.67
D	0.33
V	0
J	0
A	0
P	0



"Girls are falling"

If Sequence is StartState → Noun

P(are , Verb)

= P(are| Verb) * P(Verb | Noun) * **P(Girls | Noun)** * **P(Noun | StartState)**

= 0.43 * 0.63 * **0.1675**

= 0.04537

P(are, Noun) = P(are| Noun) * P(Noun | Noun) * **~~Noun~~** = 0 * 0 = 0

P(are, Determiner) = P(are, Adverb) = P(are , Adjective) =P(are , Preposition/Conjunction) = 0

StartState→ Noun→ Verb

If Sequence is StartState → Verb

P(are , Verb)

= P(are| Verb) * P(Verb | Verb) * **P(Girls | Verb)** *

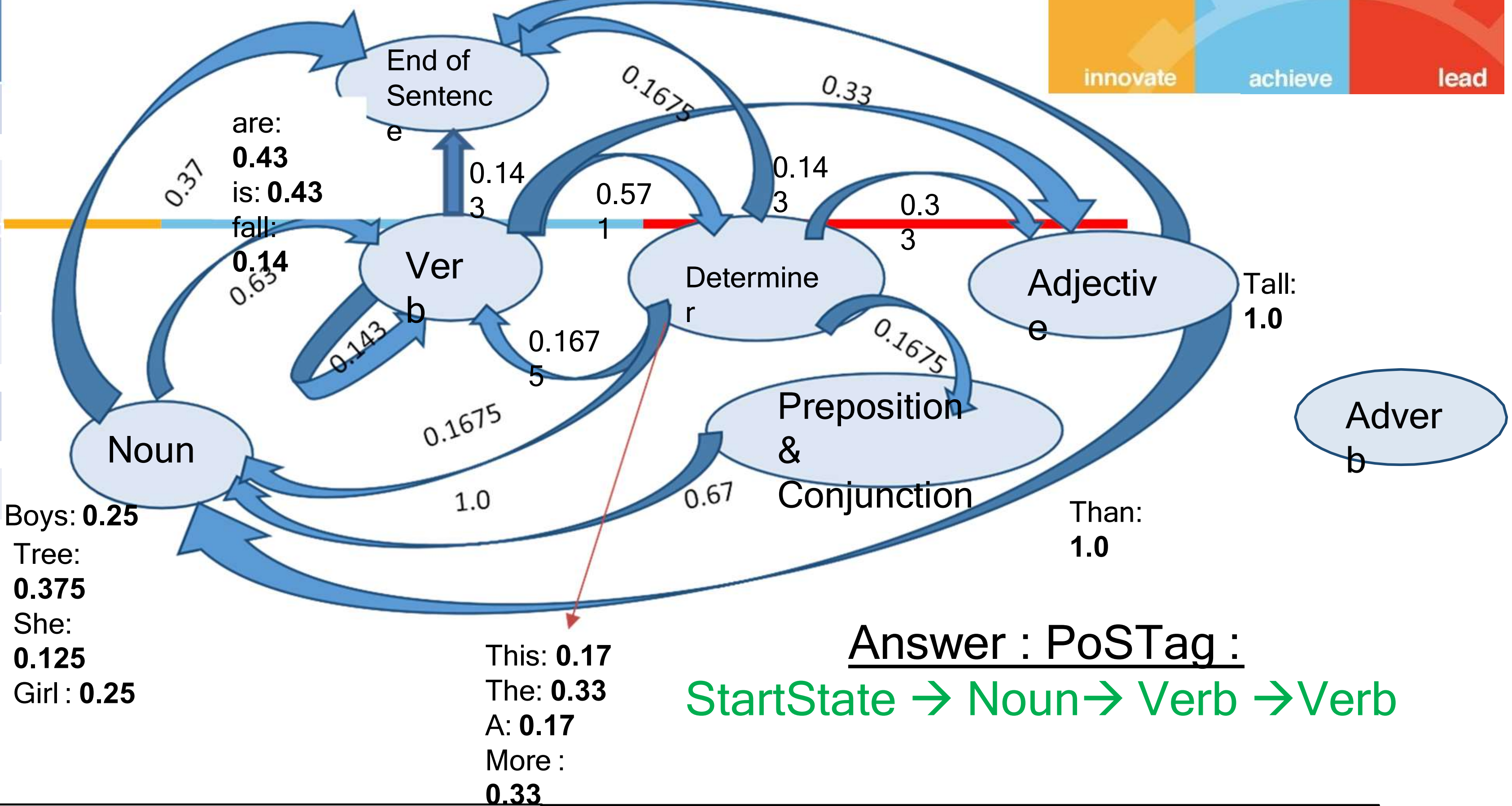
P(Verb | StartState)

= 0.43 * 0.143 * **0 = 0**

P(are, Noun) = P(are| Noun) * P(Noun | Noun) * **Verb**= 0 * 0 = 0

P(are, Determiner) = P(are, Adverb) = P(are , Adjective) =P(are , Preposition/Conjunction) = 0

Initi	
N	0.67
D	0.33
V	0
J	0
A	0
P	0



"Girls are falling"

Answer : PoSTag :

StartState → Noun→ Verb →Verb

If Sequence is StartState→ Noun →Verb

P(falling, Verb)

= P(falling|Verb)*P(Verb|Verb)***P(are| Verb)** * **P(Verb | Noun)** * **P(Girls | Noun)** * **P(Noun | StartState)**

= 0.14 * 0.143 * **0.04537**

= 0.000908

P(are, Noun) = P(are| Noun) * P(Noun | Noun) * **Noun** = 0 * 0 = 0

P(are, Determiner) = P(are, Adverb) = P(are , Adjective) =P(are , Preposition/Conjunction) = 0

If Sequence is StartState→ Verb ->Adjective

P(falling , Verb)

= P(falling | Verb) * P(Verb | Verb) * P(are| Adjective) * P(Adjective | Verb) * P(Girls | Verb) * P(Verb | StartState)

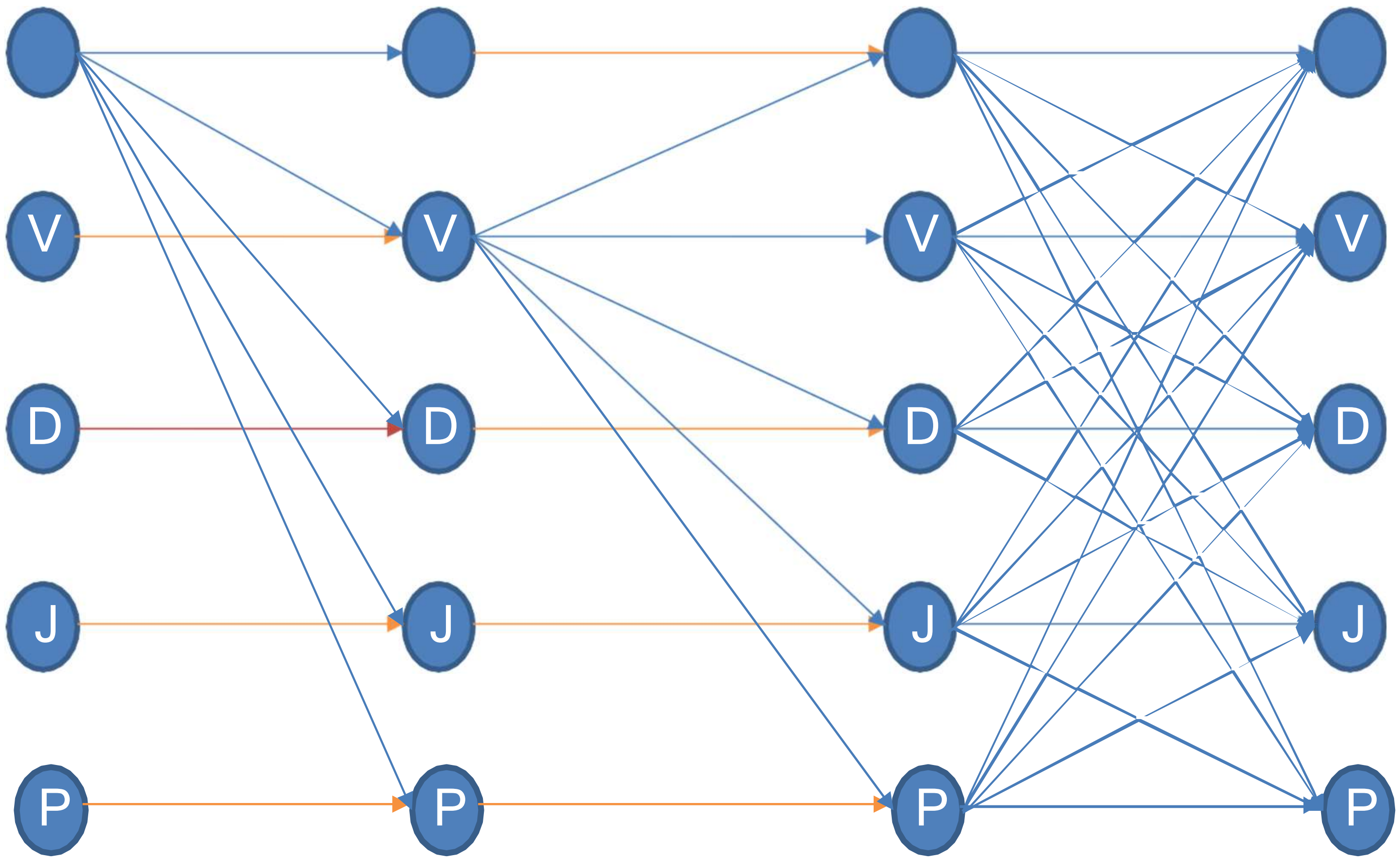
= 0.14 * 0.143 * **0 = 0**

P(falling, Noun) = P(falling| Noun) * P(Noun | Adjective) * **Verb**= 0 * 0 = 0

P(falling, Determiner) = P(falling, Adverb) = P(falling , Adjective) =P(falling , Preposition/Conjunction) = 0

Sample Sequence under Test: Start -> Noun -> Verb -> … … … .

Assume Noun -> Verb is the maximum Value



Learning HMM Parameters

Parameter Estimation by EM Algorithm

(Baum-Welch re-estimation procedure)

Parameter Estimation

Learning Approach

Baum-Welch re-estimation procedure: Backward Propagation Algorithm

Given an observation sequence O(Evidence) and the set of possible states in the HMM, learn the HMM parameters A(Transition) and B(Emission).

Given set of weather observations recorded estimate the PARAMETERS:

{SS, SR, RR}

	HH	HL	LH	LL
SS	**(0.5).(0.6).(0.8)(0.6) = 0.1440**	0.0120	0.03	0.01
SR	**0.0960**	0.048	0.02	0.04
RR	0.064	0.032	0.12	**0.16**
Total	0.304	0.092	0.17	0.21

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probabi

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Parameter Estimation

Learning Approach

Baum-Welch re-estimation procedure: Backward Propagation Algorithm

Given an observation sequence O(Evidence) and the set of possible states in the HMM, learn the HMM parameters A(Transition) and B(Emission).

Given set of weather observations recorded estimate the PARAMETERS:

{SS, SR, RR}

	HH	HL	LH	LL	Best Seq	P(Best)
SS	**0.1440**	0.0120	0.03	0.01	HH	0.144
SR	**0.0960**	0.048	0.02	0.04	HH	0.096
RR	0.064	0.032	0.12	**0.16**	LL	0.16
Total	0.304	0.092	0.17	0.21		0.4
Normalize	0.76	0.23	0.425	0.525		

HP	LP	
0.232323232	0.5526316	LP
0.767676768	0.4473684	HP

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probabi

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Parameter Estimation

Learning Approach

Baum-Welch re-estimation procedure: Backward Propagation Algorithm

Given an observation sequence O(Evidence) and the set of possible states in the HMM, learn the HMM parameters A(Transition) and B(Emission).

Find set of weather observations recorded estimate the parameters:

{SS, SR, RR}

After this step for the second iteration
Use the optimized tables
(Initial, Transition , Emission)
and repeat the algorithm till convergence

	Start(H)	Start(L)	Best Seq	P(Best)
SS	**0.1440**	0.03	HH	0.144
SR	0.0960	0.04	HH	0.096
RR	0.064	0.16	LL	0.16
	0.304	0.23		
Normalize	0.76	0.575		

HP	LP
0.56929	0.4307

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Parameter Estimation

Learning Approach

Baum-Welch re-estimation procedure: Backward Propagation Algorithm

Given an observation sequence O(Evidence) and the set of possible states in the HMM, learn the HMM parameters A(Transition) and B(Emission).

Find set of weather observations recorded estimate the parameters:

{SS, SR, RR}

LP	HP	
0.615384615	0.4	R
0.384615385	0.6	S

	H→S	L→S	H→R	L→R	Best Seq	P(Seq)
SS	**0.1440**	0.01			HH	0.144
SR	0.0960	0.04	0.096	0.048	HH	0.096
RR			0.064	0.0320	LL	0.16
Total	0.24	0.05	0.16	0.08		
Normalize	0.6	0.125	0.4	0.2		

Transition Model / Probability Matrix

$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probabi

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

HMM in Prevention of Network Security Threat

(Interesting Case Studies)

Hidden Morkov Model

Cyber Security

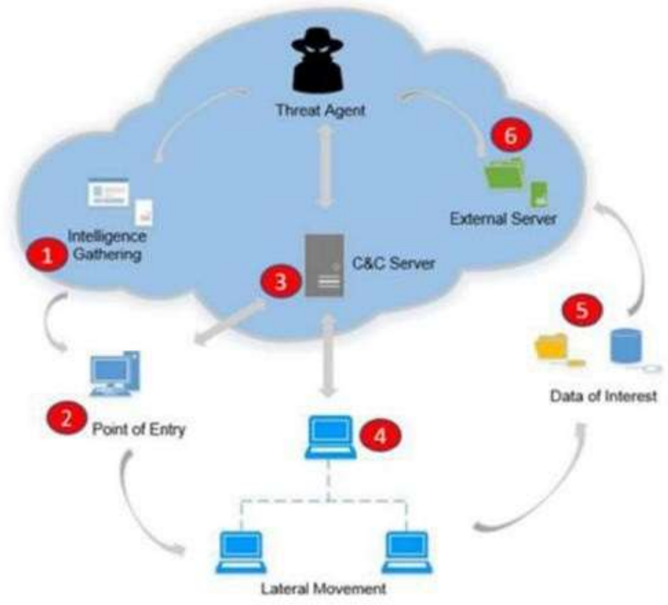
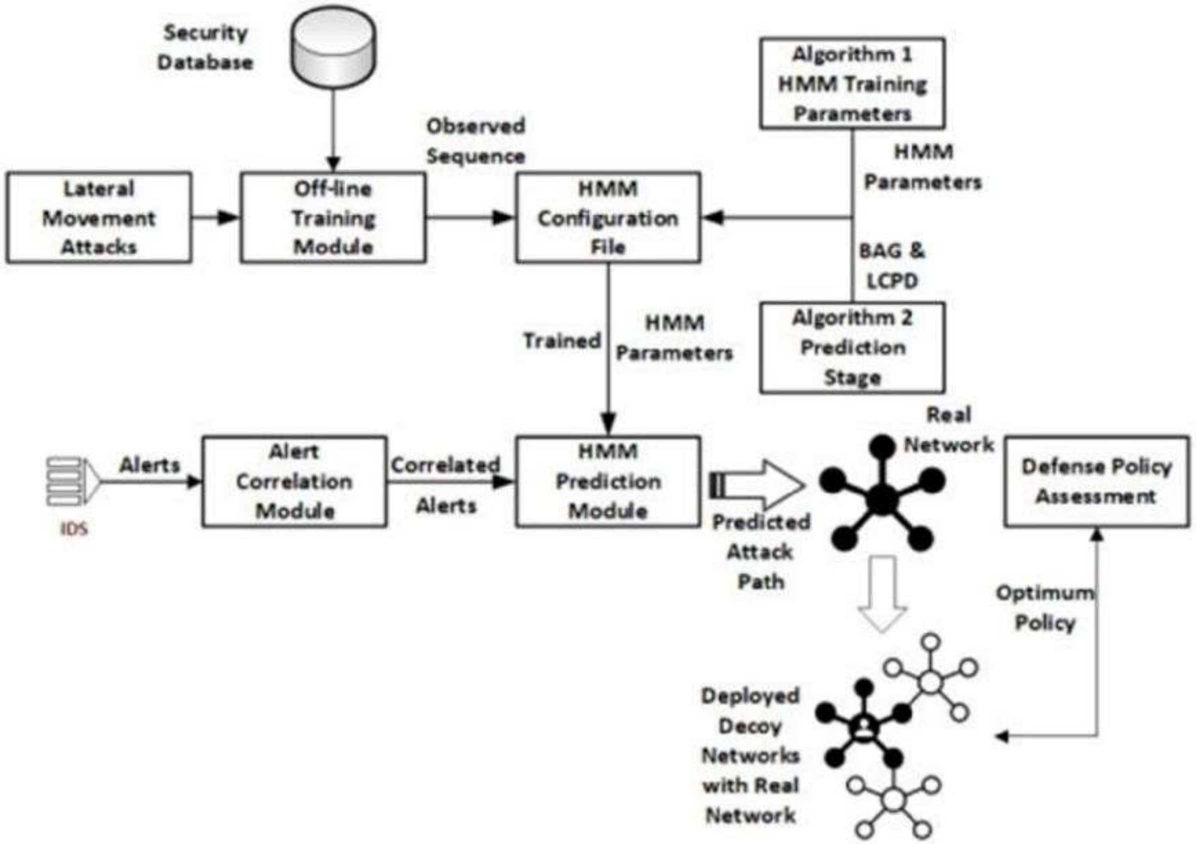


FIGURE 1. **Typical stages of APT attack.**

Source Credit : [2021 : Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement]

Hidden Morkov Model

Cyber Security



Source Credit : [2021 : Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement]

Hidden Morkov Model

Cyber Security

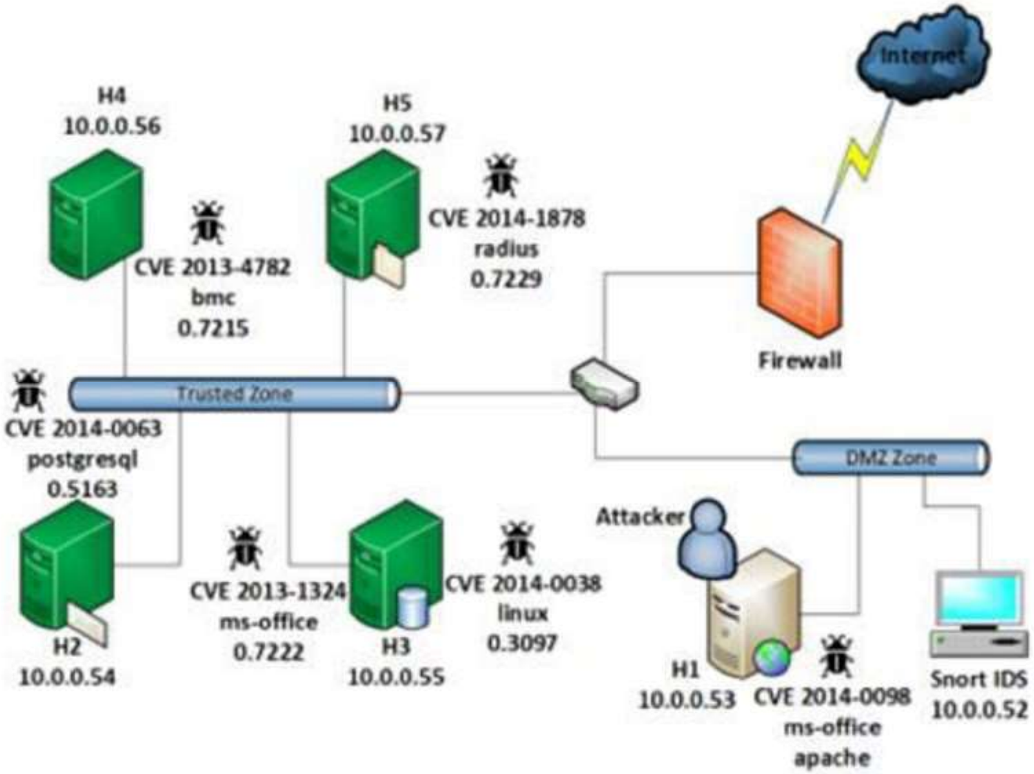


FIGURE 9. Experimental network topology.

Source Credit : [2021 : Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement]

Hidden Morkov Model

Cyber Security

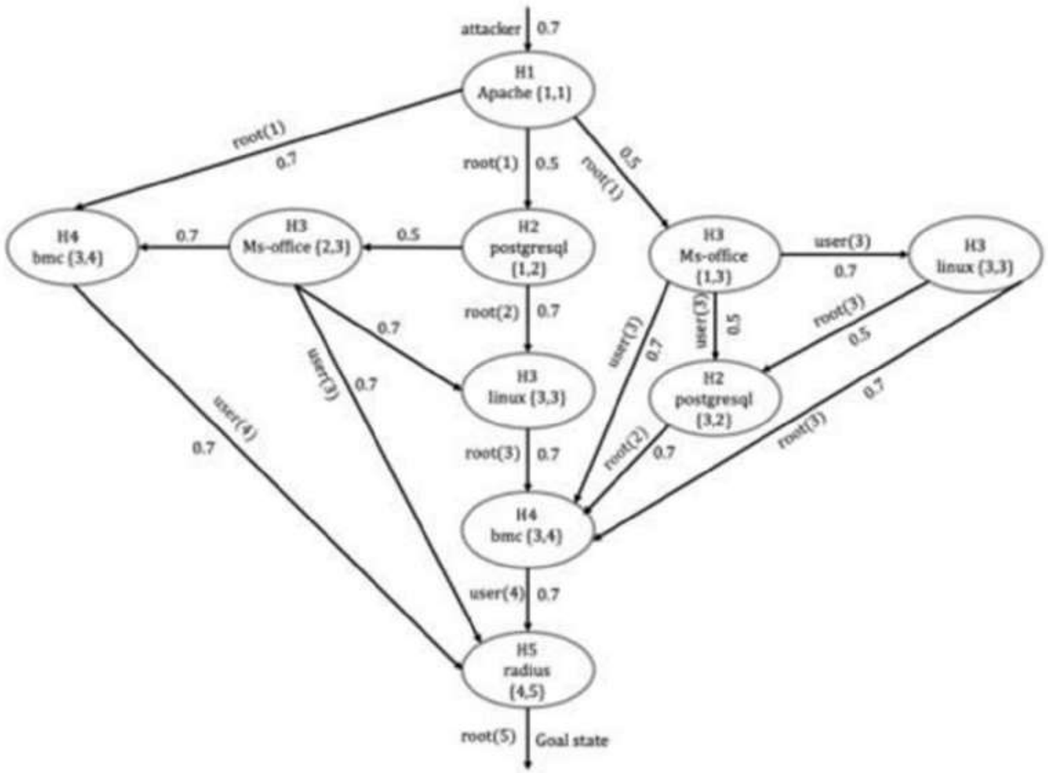


FIGURE 10. Attack graph of the experimental network.

Attack states description.

State	Description
S_1	Initial State
S_2	$(H_1$,root)
S_3	$(H_2$,root)
S_4	$(H_3$,user)
S_5	$(H_3$,root)
S_6	$(H_4$,user)
S_7	$(H_5$,root)

Source Credit : [2021 : Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement]

Hidden Morkov Model

Cyber Security

Attack states description.

State	Description
S_1	Initial State
S_2	$(H_1$,root)
S_3	$(H_2$,root)
S_4	$(H_3$,user)
S_5	$(H_3$,root)
S_6	$(H_4$,user)
S_7	$(H_5$,root)

TABLE 6. Possible attack paths.

Path Number	Attack Path
1	$S_1 \rightarrow S_2 \rightarrow S_6 \rightarrow S_7$
2	$S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_7$
3	$S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_6 \rightarrow S_7$
4	$S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_5 \rightarrow S_6 \rightarrow S_7$
5	$S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_5 \rightarrow S_6 \rightarrow S_7$
6	$S_1 \rightarrow S_2 \rightarrow S_4 \rightarrow S_6 \rightarrow S_7$
7	$S_1 \rightarrow S_2 \rightarrow S_4 \rightarrow S_3 \rightarrow S_6 \rightarrow S_7$
8	$S_1 \rightarrow S_2 \rightarrow S_4 \rightarrow S_5 \rightarrow S_3 \rightarrow S_6 \rightarrow S_7$
9	$S_1 \rightarrow S_2 \rightarrow S_4 \rightarrow S_5 \rightarrow S_6 \rightarrow S_7$

Source Credit : 2021 : Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement

Required Reading: AIMA - Chapter #15.1, #15.2, #15.3, # 20.3

Thank You for all your Attention

Note : Some of the slides are adopted from AIMA TB materials