Artificial & Computational Intelligence

AIMLCZG557

Contributors & Designers of document content : Cluster Course Faculty Team

M6 : Reasoning over time

BITS Pilani
Pilani Campus

Presented by
Faculty Name
BITS Email ID

Artificial and Computational Intelligence

Disclaimer and Acknowledgement

- Few content for these slides may have been obtained from prescribed books and various other source on the Internet
- I hereby acknowledge all the contributors for their material and inputs and gratefully acknowledge people others who made their course materials freely available online.
- .I have provided source information wherever necessary
- This is not a full fledged reading materials. Students are requested to refer to the textbook w.r.t detailed content of the presentation deck that is expected to be shared over e-learning portal - taxilla.
- I have added and modified the content to suit the requirements of the class dynamics & live session's lecture delivery flow for presentation

- **Slide Source / Preparation / Review:**
- From BITS Pilani WILP**:** Prof.Raja vadhana, Prof. Indumathi, Prof.Sangeetha
- From BITS Oncampus & External : Mr.Santosh GSK

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

Learning Objective

1. Understand the relationship between Time & Uncertainty

1. Recognize the transition model of Markov Model

1. Relate to the application of the Hidden Markov Model

Module 6:
Reasoning over time

Reasoning Over Time

A. Time and Uncertainty

B. Inference in temporal models

C. Introduction to Hidden Markov Model

D. Applications of HMM

Sequential Decision Problems & Markov Decision Process

Markov Decision Process

Sequential Problem | Partial Observability | Belief System

Modelling sequences of random events and transitions between states over time is known as Markov chain

Agents in partially observable environment should keep a track of current state to the extent allowed by sensors

E.g., Robot moving in a new maze

Agent maintains a **belief state** representing the current possible world states

Transition Model / Probability Matrix :

Using belief state and transition model,
the agent can how the world might evolve in
next time step. To capture the degree of belief
we will use Probability Theory. We model the change in world
using a variable for each aspect of state and
at each point in time.
Current state depends only finite number of previous states.

$$\begin{array}{cc} C & M \\ \left|\begin{array}{cc} 0.40 & 0.20 \\ 0.60 & 0.80 \end{array}\right| & \begin{array}{c} C \\ M \end{array} \end{array}$$

Markov Decision Process

Time - Uncertainty | States - Observations

Static World: Each random variable would have a single fixed value

E.g., Diagnosing a broken car

Dynamic World: The state information keeps changing with time

E.g., treating a diabetic patient, tracking the location of robot, tracking economic activity of a nation

Time slices: World is observed in time slices. Each slice has a set of random variables, some observable and some not.

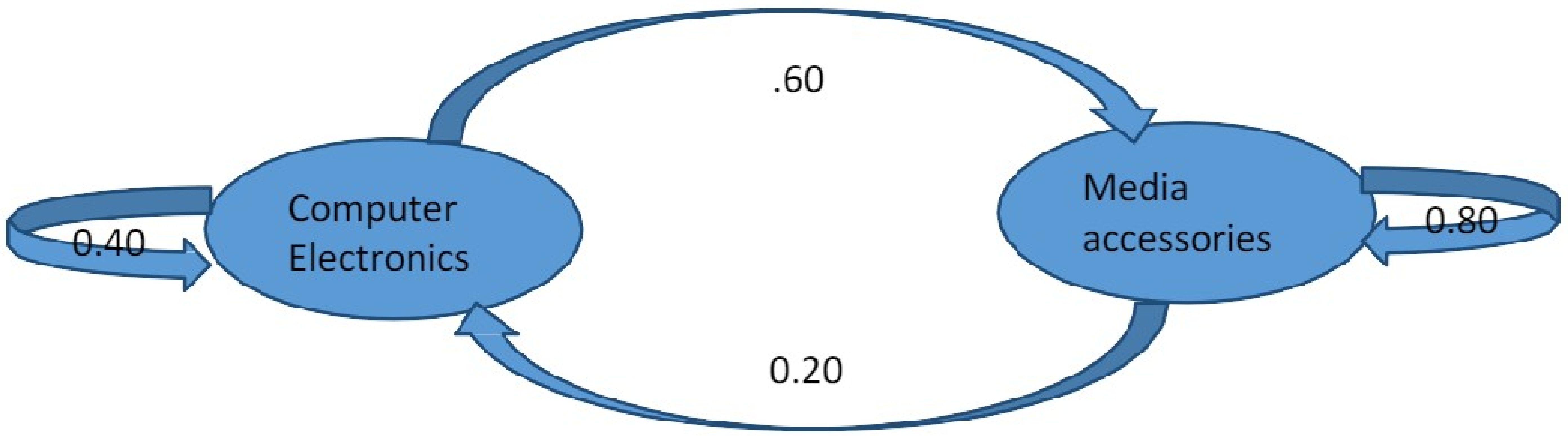
Assumption: We will assume same subset of random variables are observable in each time slice

$\mathbf{E_t}$ - set of observable random variables at time t

$\mathbf{X_t}$ - set of unobserved random variables at time t

$$\begin{array}{cc} \mathbf{C} & \mathbf{M} & \\ \left|\begin{matrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{matrix}\right| & & \begin{matrix} \mathbf{C} \\ \mathbf{M} \end{matrix} \end{array}$$

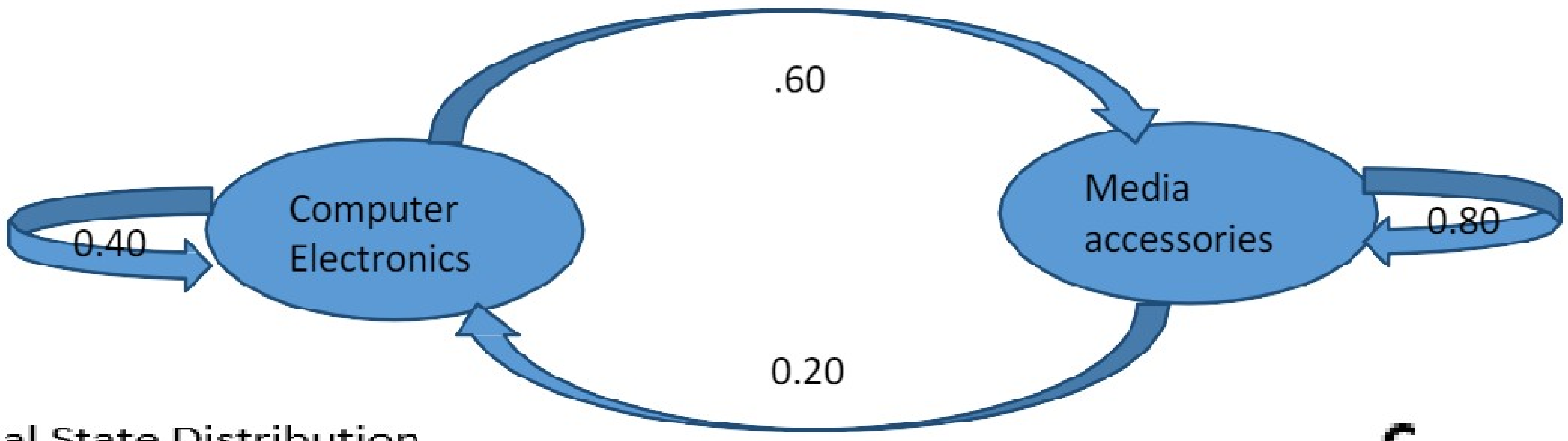
Markov Model- Example 1



Transition Model

	C	M	
	0.40	0.20	C
	0.60	0.80	M

Markov Model



Current State: Initial State Distribution

$$\begin{array}{c} \begin{vmatrix} 1 \\ 0 \end{vmatrix} \end{array} \begin{array}{l} \mathbf{C} \\ \mathbf{M} \end{array}$$

$$\begin{array}{cc} \mathbf{C} & \mathbf{M} \\ \end{array}$$
$$\begin{vmatrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{vmatrix} \begin{array}{l} \mathbf{C} \\ \mathbf{M} \end{array}$$

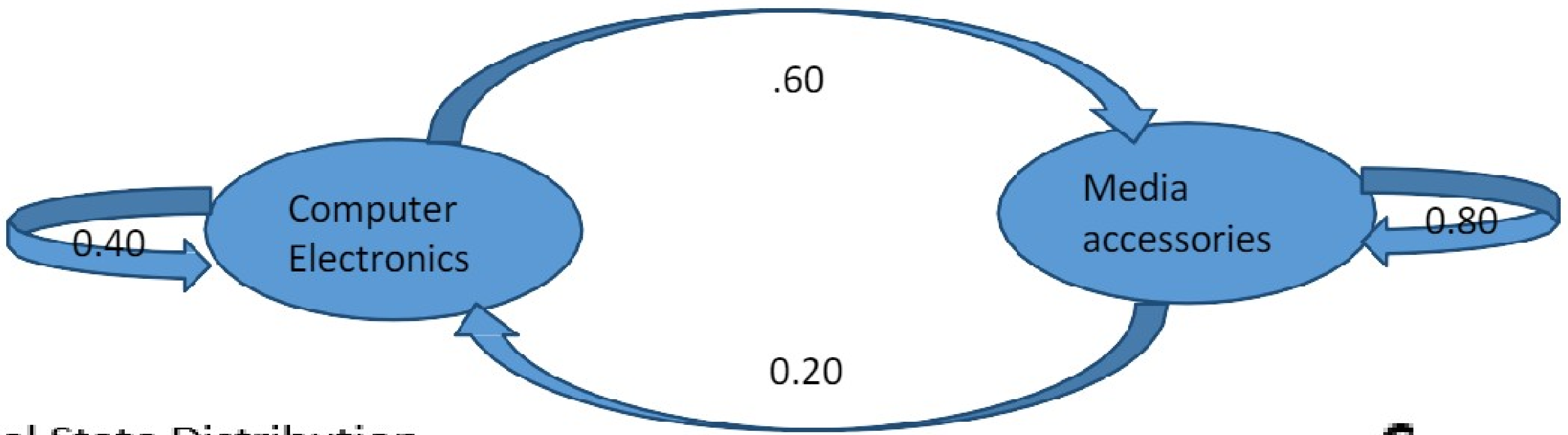
Next State : Likely to buy Media accessories on next visit

$$\begin{vmatrix} 0.40 \\ 0.60 \end{vmatrix} \begin{array}{l} \mathbf{C} \\ \mathbf{M} \end{array}$$

Next State : Likely to buy Media accessories on next visit

$$\begin{vmatrix} 0.28 \\ 0.72 \end{vmatrix} \begin{array}{l} \mathbf{C} \\ \mathbf{M} \end{array}$$

Markov Model



Current State: Initial State Distribution

$$\begin{array}{c|c} \left|\begin{matrix} 1 \\ 0 \end{matrix}\right| & \begin{matrix} \mathbf{C} \\ \mathbf{M} \end{matrix} \end{array}$$

Next State : Likely to buy Media accessories on next visit

$$\begin{array}{c|c} \left|\begin{matrix} 0.40 \\ 0.60 \end{matrix}\right| & \begin{matrix} \mathbf{C} \\ \mathbf{M} \end{matrix} \end{array}$$

Next State : Likely to buy Media accessories on next visit

$$\begin{array}{c|c} \left|\begin{matrix} 0.28 \\ 0.72 \end{matrix}\right| & \begin{matrix} \mathbf{C} \\ \mathbf{M} \end{matrix} \end{array}$$

$$\begin{array}{cc} \begin{matrix} \mathbf{C} & \quad\quad & \mathbf{M} \end{matrix} & \\ \left|\begin{matrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{matrix}\right| & \begin{matrix} \mathbf{C} \\ \mathbf{M} \end{matrix} \end{array}$$

Markov Process

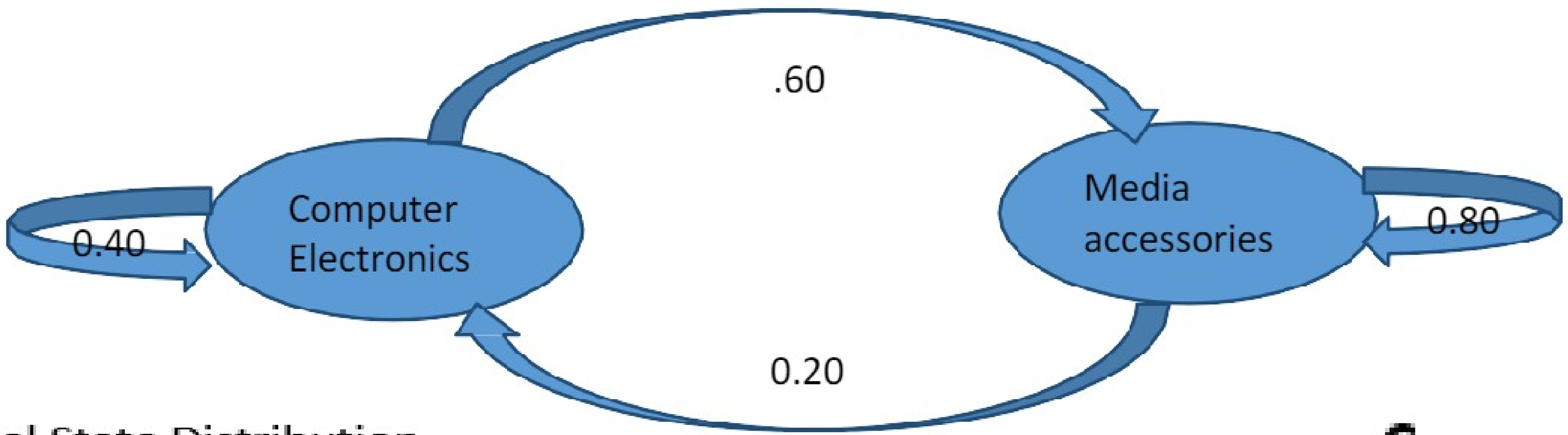
States | Observations | Assumptions

Modelling sequences of random events and transitions between states over time is known as Markov chain

Transition Model / Probability Matrix :
Current state depends only finite number of previous states. :

$$
\begin{array}{cc}
\begin{array}{cc} \textbf{C} & \textbf{M} \end{array} & \\
\left|\begin{array}{cc} 0.40 & 0.20 \\ 0.60 & 0.80 \end{array}\right| & \begin{array}{c} \textbf{C} \\ \textbf{M} \end{array}
\end{array}
$$

Markov Model



Current State: Initial State Distribution

$$\begin{array}{c|c}
\begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{array}{c} C \\ M \end{array}
\end{array}$$

$$\begin{array}{cc|c}
C & M & \\
\hline
0.40 & 0.20 & C \\
0.60 & 0.80 & M
\end{array}$$

Next State : Likely to buy Media accessories on next visit

$$\begin{array}{c|c}
\begin{bmatrix} 0.40 \\ 0.60 \end{bmatrix} & \begin{array}{c} C \\ M \end{array}
\end{array}$$

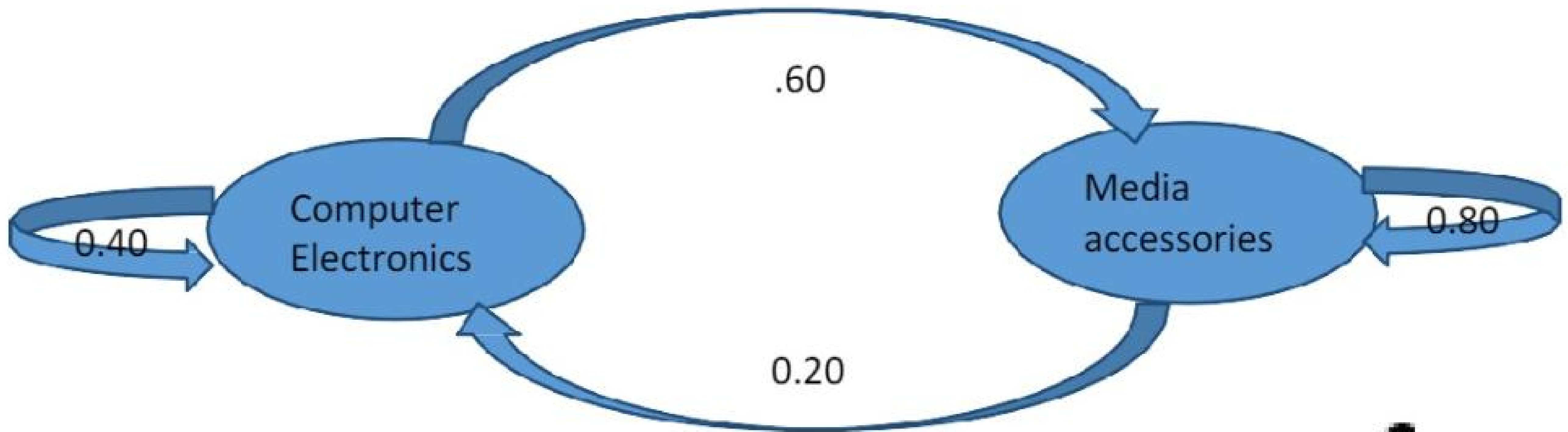
Next State : Likely to buy Media accessories on next visit

$$\begin{array}{c|c}
\begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix} & \begin{array}{c} C \\ M \end{array}
\end{array}$$

Inference in temporal Models

Markov Model

Inference Type 1



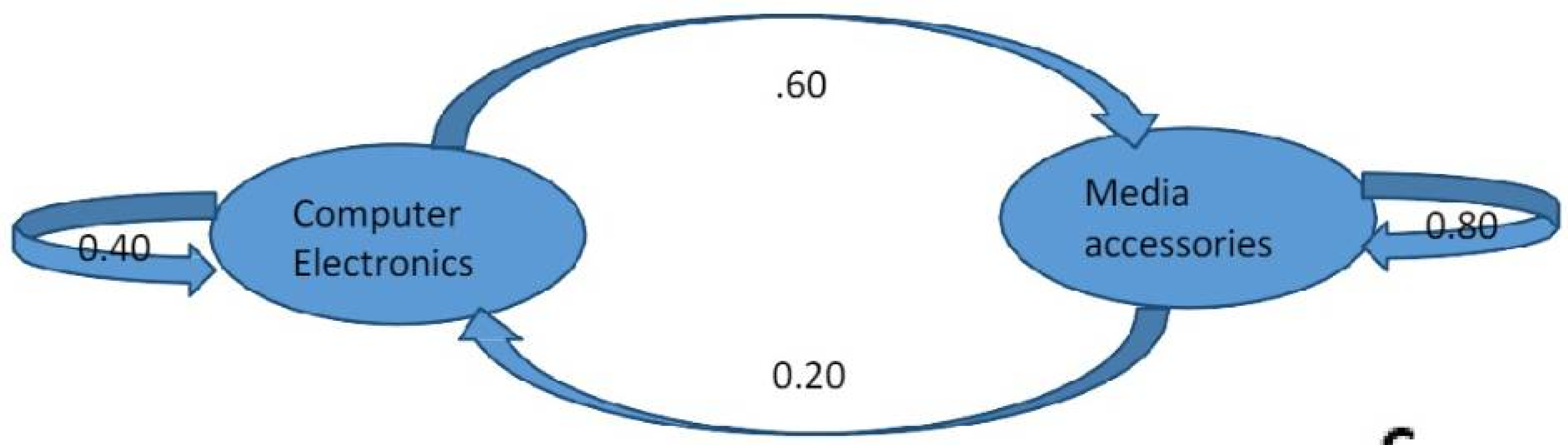
C	M	
0.40	0.20	C
0.60	0.80	M

What is the probability that the purchasing behaviour of the customer is in below sequential order only?Initial Probability Matrix is P(C) = 1 , P(M) = 0
(Computer, Media, Media, Computer)

Apply Bayes chain rule:
P(Computer, Media, Media, Computer) = P(C) * P(M|C) * P(M|M) * P(C|M) = 0.096

Markov Model

Inference Type 2



$$\begin{array}{cc|c} C & M & \\ \hline 0.40 & 0.20 & C \\ 0.60 & 0.80 & M \end{array}$$

What is the probability that the customer who purchased Media accessories will kee
coming back to purchase media accessories in the next 2 consecutive visits only?

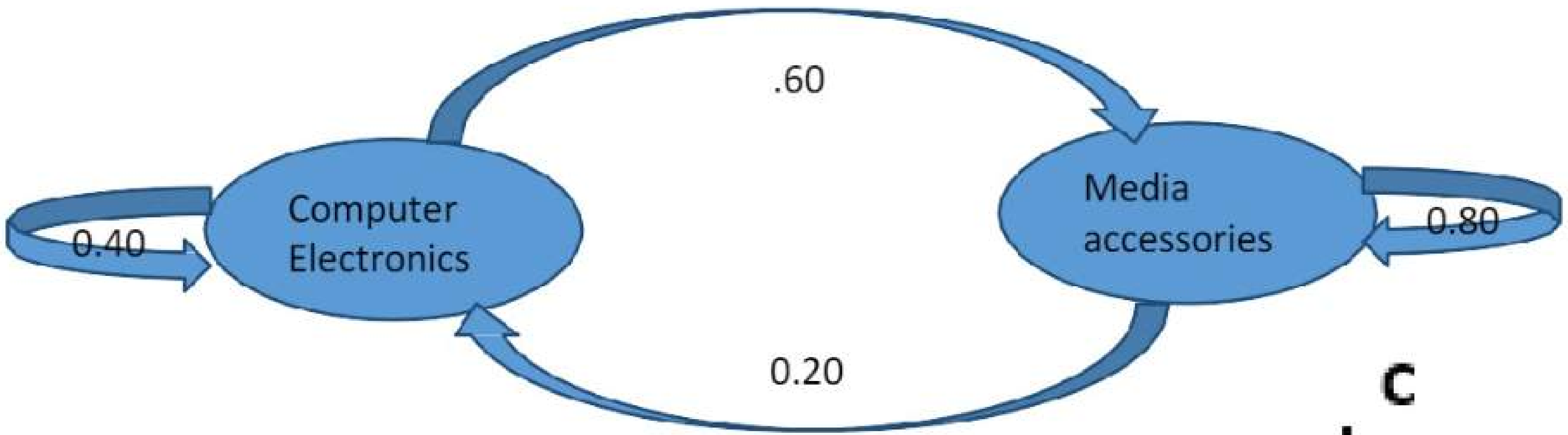
<u>Derive Initial prob values & Apply Bayes chain rule on the pattern exhibited:</u>

Initial Probability Matrix is P(M) = 1 , P(C) = 0

P(Media, Media, Media, Computer) = P(M) * P(M|M) * P(M|M) * P(C|M) = 0.128

Markov Model

Inference Type 3



$$
\begin{array}{cc|c}
C & M & \\
\begin{vmatrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{vmatrix} & & \begin{array}{c} C \\ M \end{array}
\end{array}
$$

Given the evidence that the customer walked into the store and bought a computer electronics, find the expected purchase pattern in the next 3 visits

<u>Derive Initial prob values & Apply Bayes chain rule and reverse predict the combination on the most likely pattern (Similar to Viterbi Algorithm):</u>

Initial Probability Matrix is P(C) = 1 , P(M) = 0

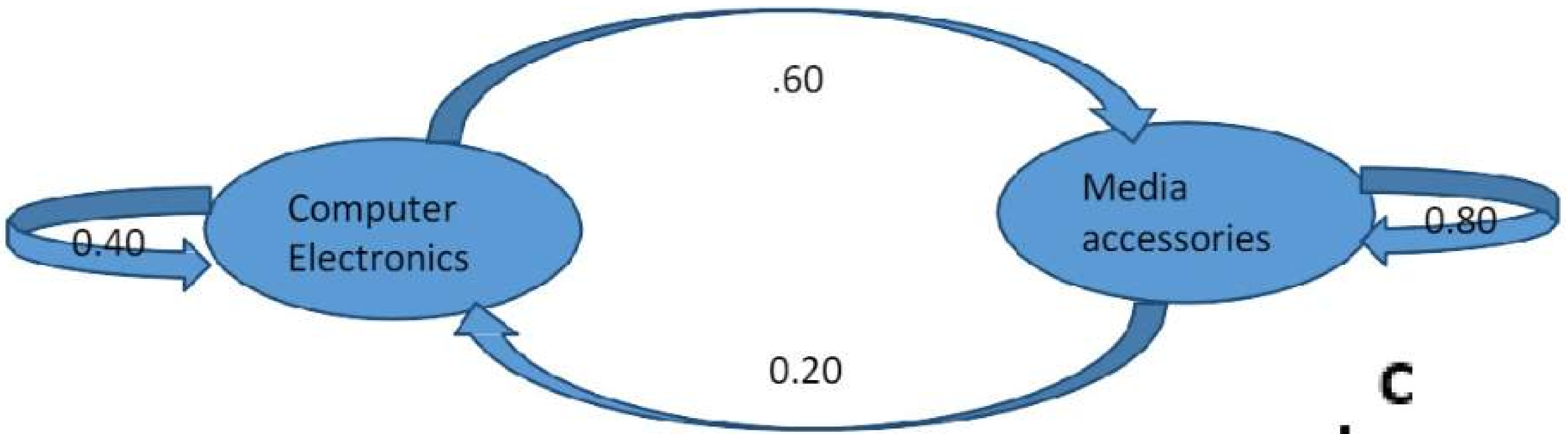
P(Computer , X, Y, Z) = P(Computer) * P(X|Computer) * P(Y|X) * P(Z|X) =

1*0.6*0.8*0.8 → Produces max values

Ans : Pattern = (Computer, Media, Media, Media)

Markov Model

Inference Type 3



$$\begin{array}{cc} C & M \\ \left|\begin{matrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{matrix}\right| \end{array}\begin{array}{c} \\ C \\ M \end{array}$$

Given the evidence that the customer walked into the store and bought a computer electronics, find the expected purchase pattern in the next 3 visits

<u>Derive Initial prob values & Apply Bayes chain rule and reverse predict the combination on the most likely pattern (Similar to Viterbi Algorithm):</u>

Initial Probability Matrix is P(C) = 1 , P(M) = 0

P(Computer , X, Y, Z) = P(Computer) * P(X|Computer) * P(Y|X) * P(Z|X) =

1*0.6*0.8*0.8 → Produces max values

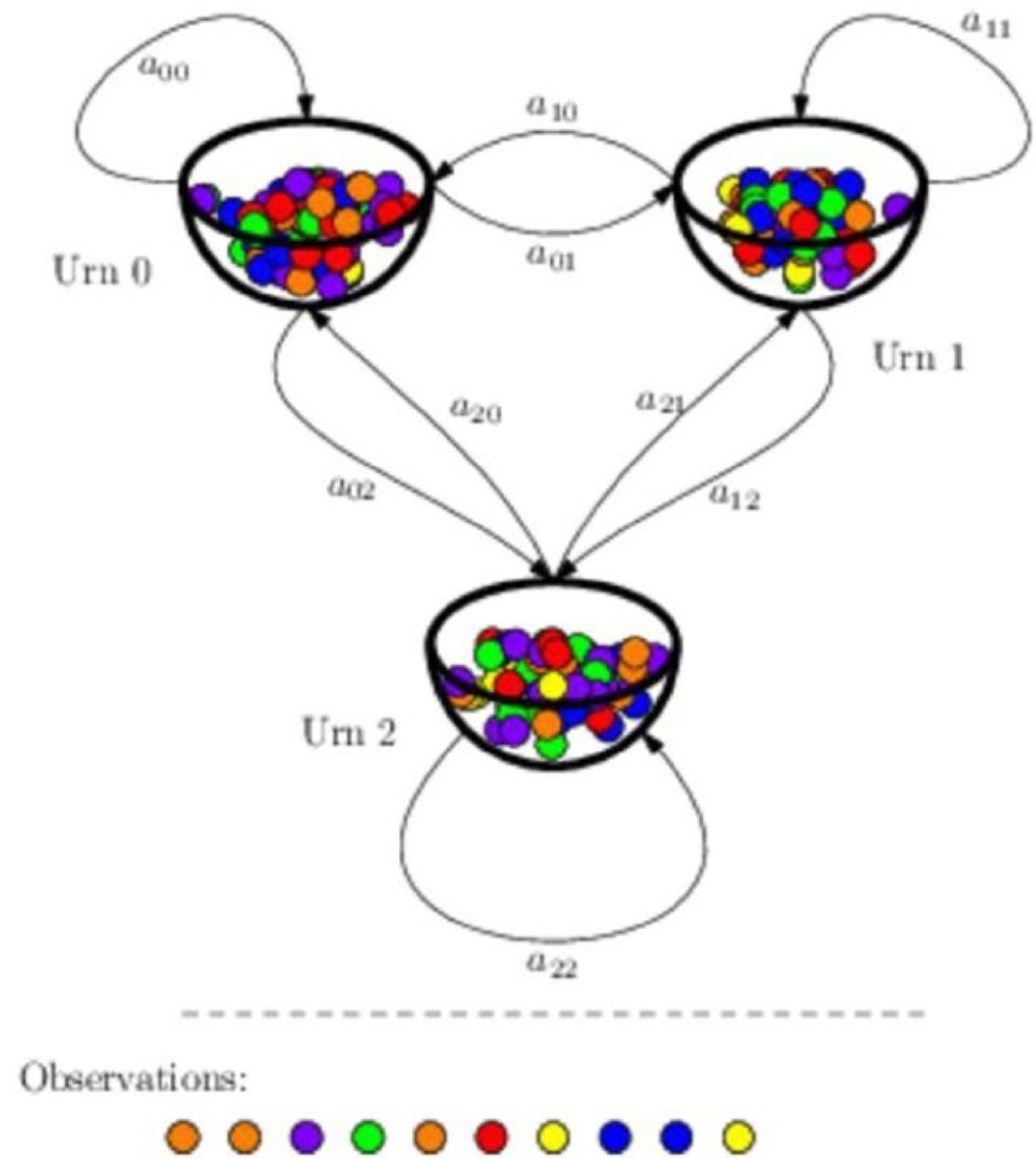
Ans : Pattern = (Computer, Media, Media, Media)

HMM

Markov Process

States | Observations | Assumptions

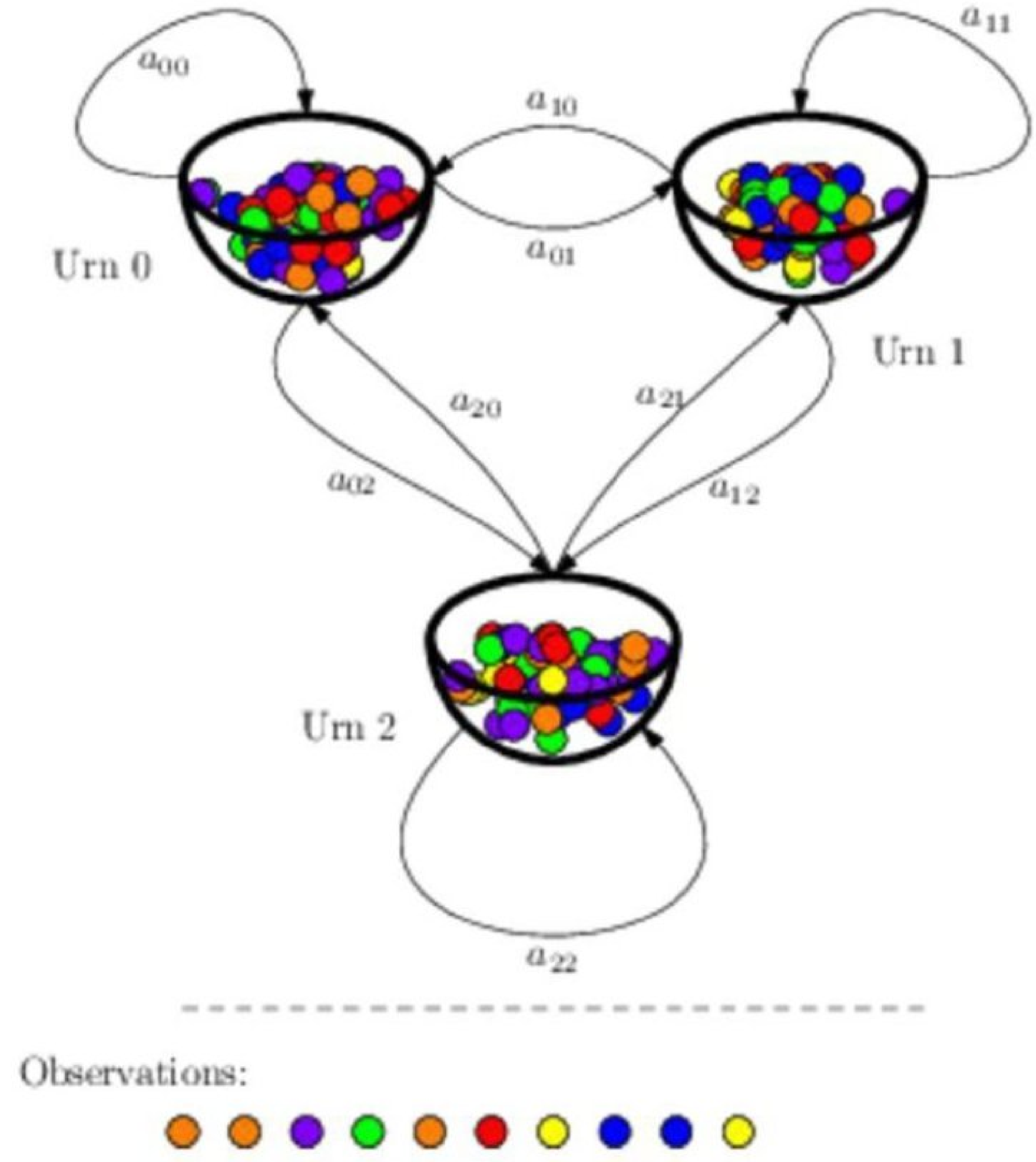
Standard Mathematical Example:
Urn & Ball Model



Markov Process

States | Observations | Assumptions

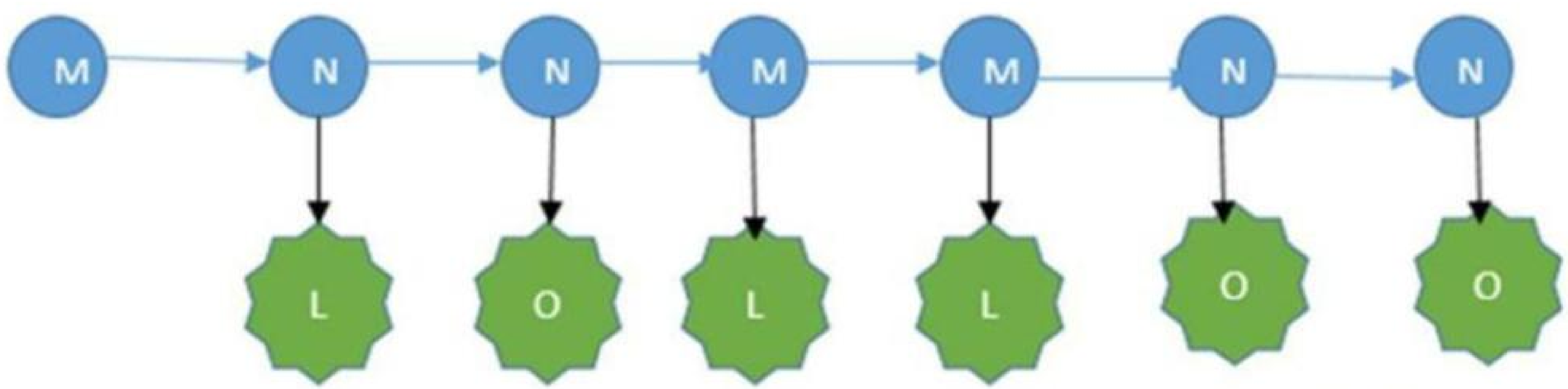
Standard Mathematical Example:
Urn & Ball Model



Hidden Markov Model

States | Observations | Assumptions

Time Slice (t)	0	1	2	3	4	5	6	$P(O_t \mid O_{t-1})$
Observed Evidence (O_t/ E_t)	-	Late	OnTime	Late	Late	Ontime	Ontime	……
Unobserved State (U_t /X_t/ Q_t)	Meeting	No Meeting	No Meeting	Meeting	Meeting	No Meeting	No Meeting	……



Transition Model / Probability Matrix

$P(U_{t-1}$ = No Meeting)	$P(U_{t-1}$ = Meeting)	←Previous
0.5	0.67	$P(U_t$ = No Meeting)
0.5	0.33	$P(U_t$ = Meeting)

Evidence / Sensor Model/ Emission Probability Matrix

$P(U_t$ = No Meeting)	$P(U_t$ = Meeting)	←Unobserved Evidence v
0.9	0.3	$P(O_t$ = OnTime)
0.1	0.7	$P(O_t$ = Late)

Hidden Markov Process

States | Observations | Assumptions

Modelling sequences of random events and transitions between states over time is known as Morkov chain

Hidden Markov Process models events as the state sequences that are not directly observable but only be approximated from the sequence of observations produced by the system

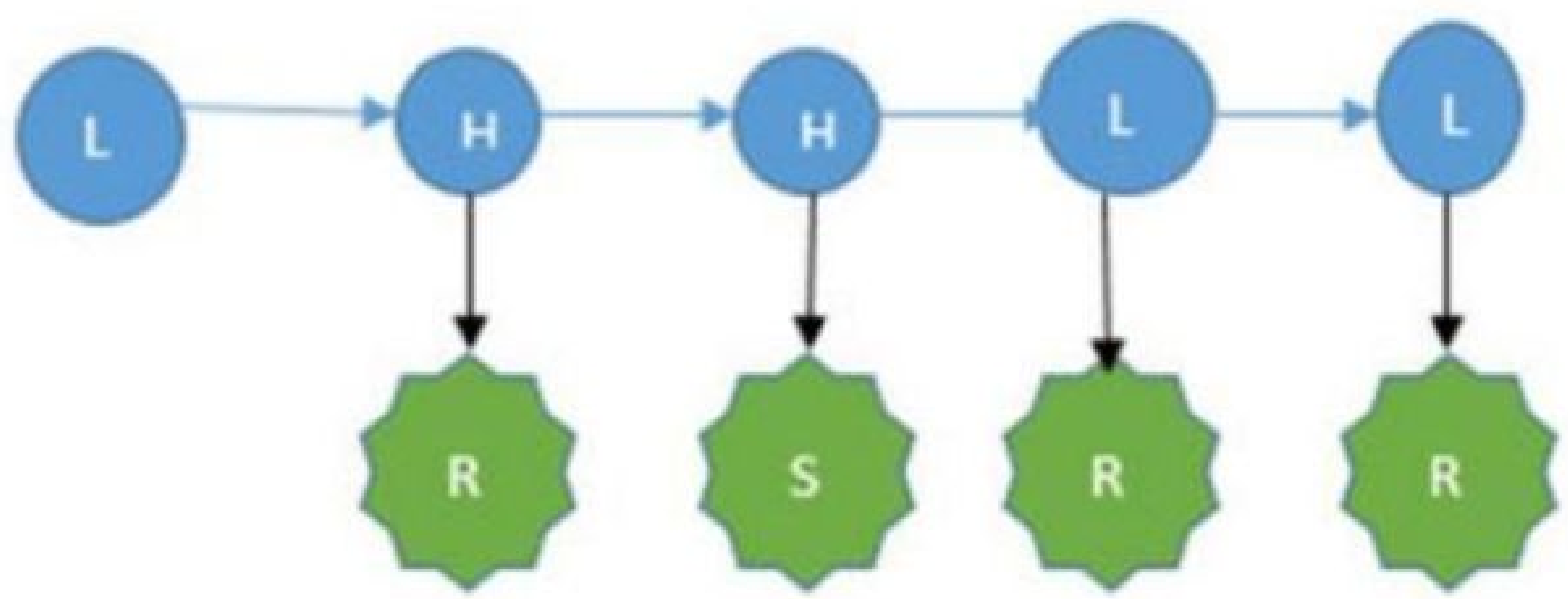
Transition Model / Probability Matrix :
Current state depends only finite number of previous states. :

Evidence / Sensor Model/ Emission Probability Matrix :
Current Evidence or Observation depends Current State of the world. Given the Current State Knowledge of the world, observation doesn't depend on history:

Hidden Morkov Model

States | Observations | Assumptions

Time Slice (t)	0	1	2	3	4	....	$P(O_t \mid O_{t-1}, O_{t-2})$
Observed Evidence (O_t)	-	Rainy	Sunny	Rainy	Rainy		
Unobserved State(U_t)	Low Pressure	High Pressure	High Pressure	Low Pressure	Low Pressure		



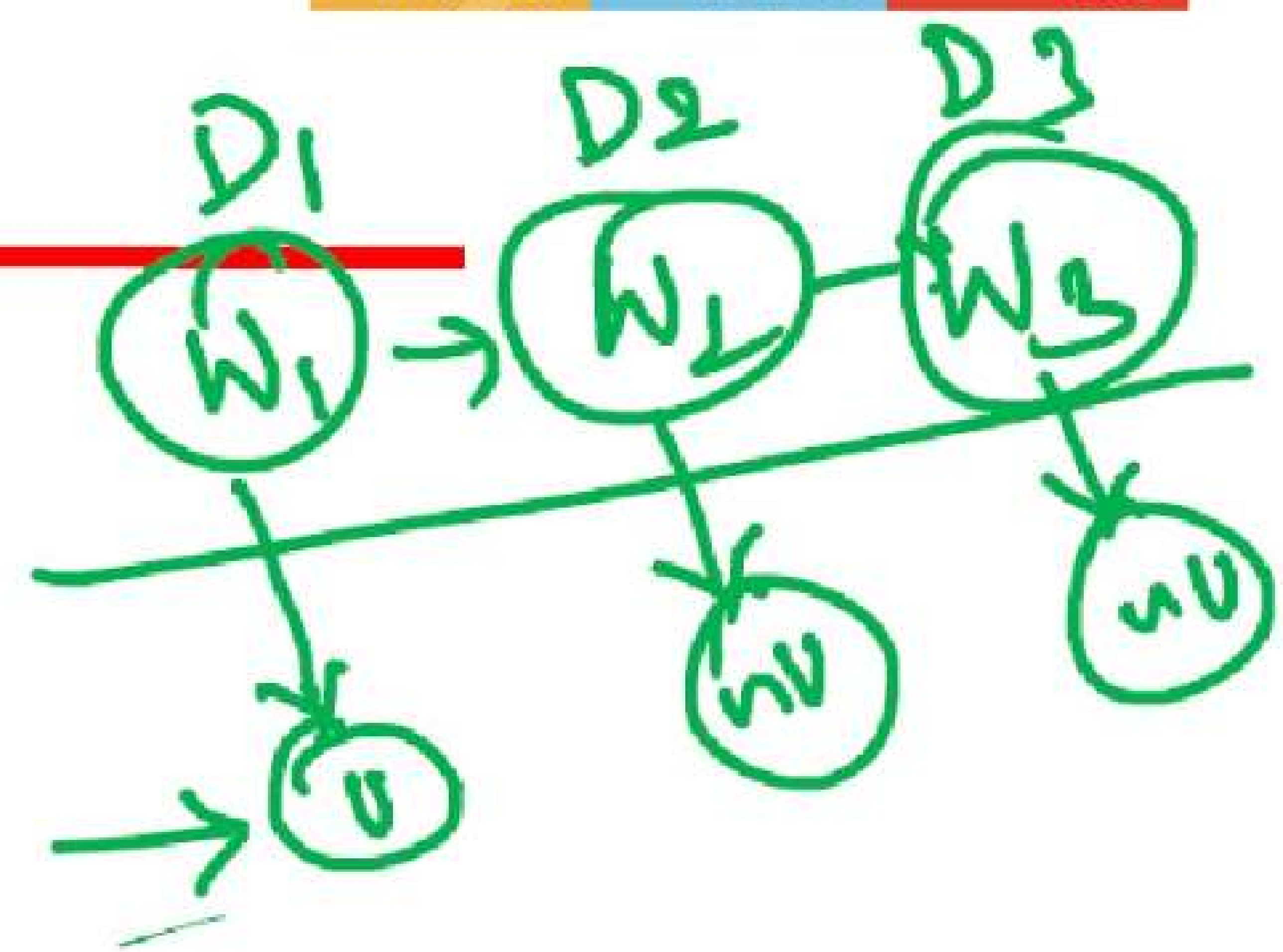
Transition Model / Probability Matrix

$P(U_{t-2} = LP, U_{t-1} = HP)$	$P(U_{t-2} = HP, U_{t-1} = HP)$	$P(U_{t-2} = HP, U_{t-1} = LP)$	$P(U_{t-2} = LP, U_{t-1} = LP)$	←Previous
0.2	0.40	0.85	0.5	$P(U_t = LP)$
0.8	0.60	0.15	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

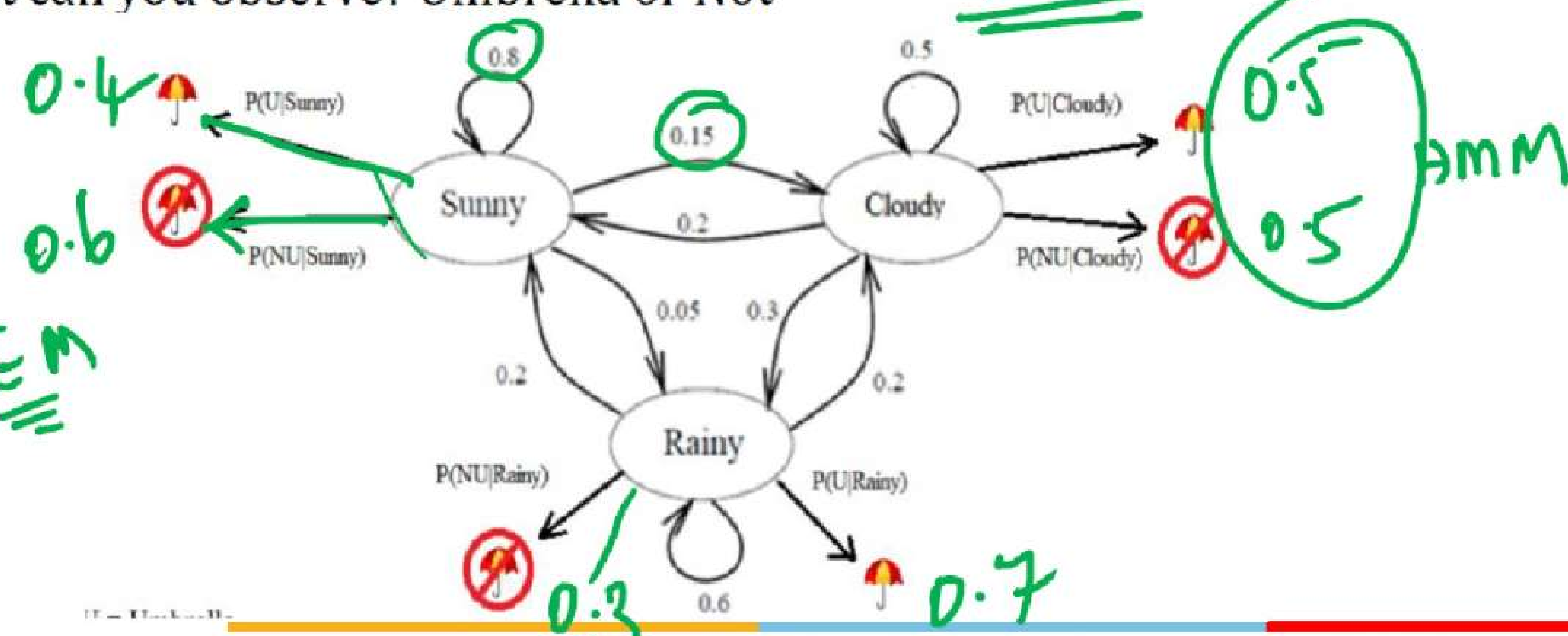
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

3rd



- Imagine: You were locked in a room for several days and you were asked about the weather outside. The only piece of evidence you have is whether the person who comes into the room bringing your daily meal is carrying an umbrella or not.
- What is hidden? Sunny, Rainy, Cloudy
- What can you observe? Umbrella or Not

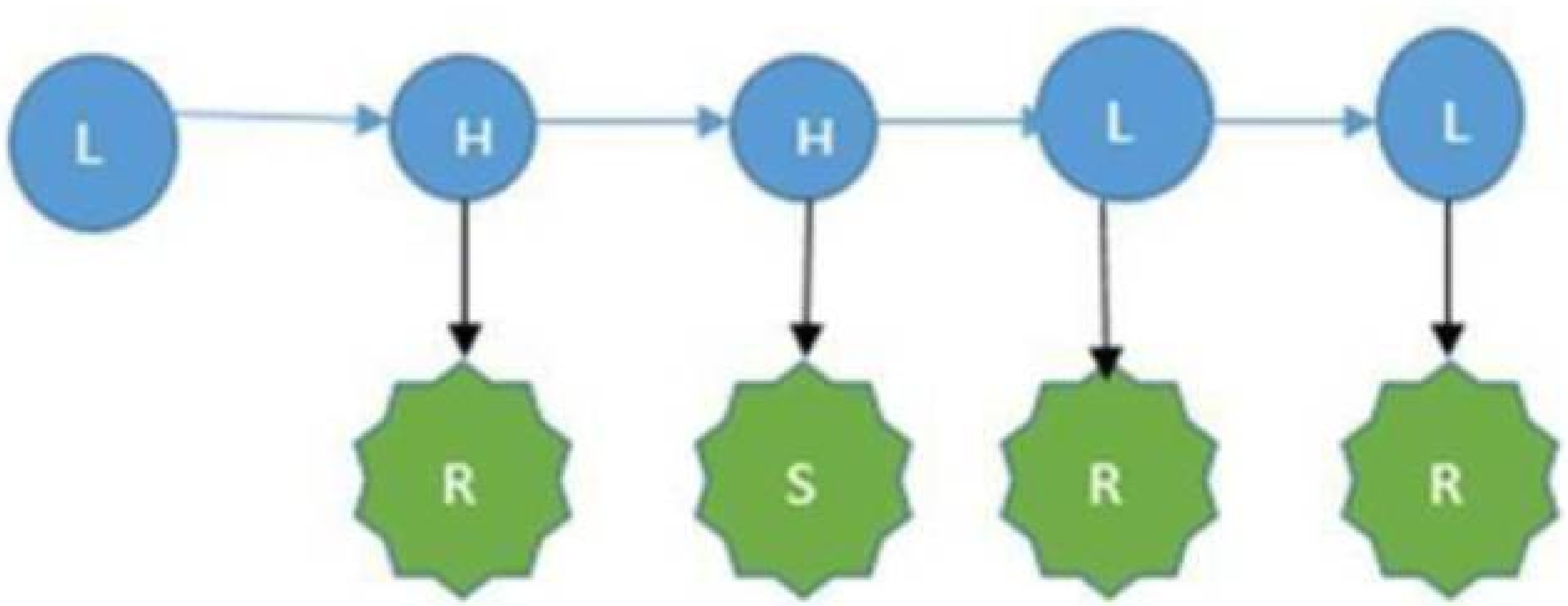
TM



Hidden Morkov Model

Filtering	Prediction	Smoothing	Most Likely Explanation
$P(L_3 \mid R\text{-}S\text{-}R\text{-}R)$	$P(L_3 \mid R\text{-}S)$	$P(H_2 \mid R\text{-}S\text{-}R\text{-}R)$	$P(H\text{-}H\text{-}L\text{-}L \mid R\text{-}S\text{-}R\text{-}R)$
$P(X_t \mid E_{1\ldots t})$	$P(X_{t+k} \mid E_{1\ldots t})$	$P(X_{k,\ o>k>t} \mid E_{1\ldots t})$	$\operatorname{argmax} X_{1\ldots t} : P(X_{1\ldots t} \mid E_{1\ldots t})$

In your Text book another example for these inferences is explained "Task of predicting the weather condition by a security personnel sitting in an underground secret installation by observing the state of an employee who either umbrella or don't" Kindly check it and work it out as additional practice

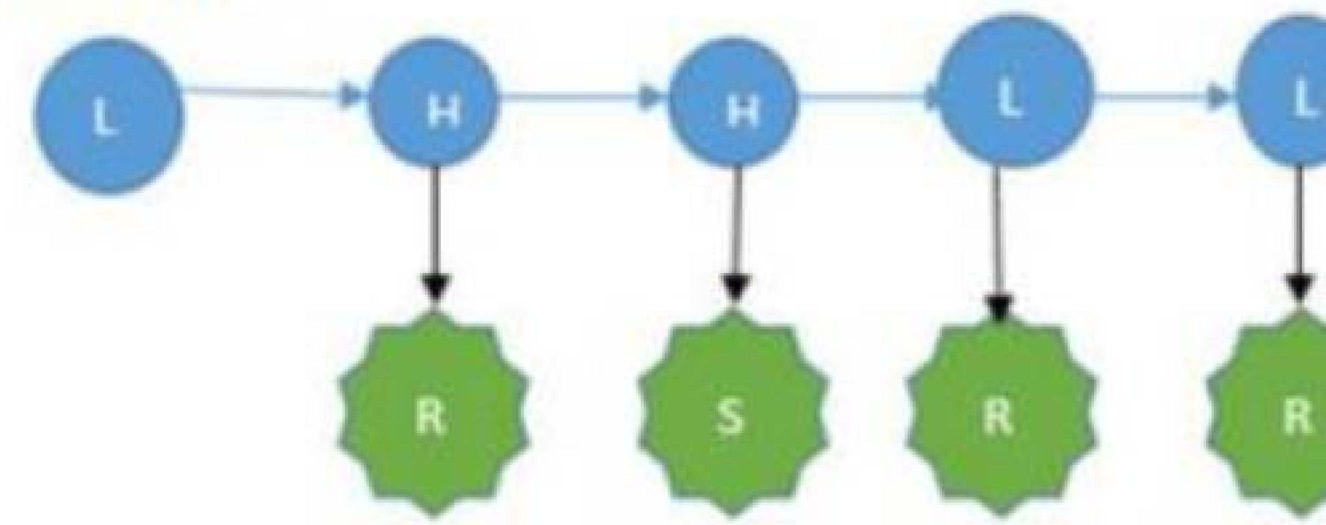


Hidden Morkov Model

Inference: Type -1

Sequence Evaluation : Likely hood Computation : Forward Algorithm

Find the probability of occurrence of
this weather sequence observation: **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	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

P(SSR)

$$= \sum_{X} P(SSR\,,\ X) = \sum_{X} P(SSR\,,\, X_1X_2X_3)$$

$$= \sum_{X} P(R, X_3, S, X_2, S, X_1) = \sum_{X} P(R \mid X_3) * P(X_3 \mid X_2) * P(S \mid X_2) * P(X_2 \mid X_1) * P(S \mid X_1) * P(X_1 \mid X_0)$$

$$= \sum_{X} P(R \mid X_3) * P(S \mid X_2) * P(S \mid X_1) * P(X_3 \mid X_2) * P(X_2 \mid X_1) * P(X_1 \mid X_0)$$

$$= \sum_{X} \prod_{j=1}^{t} P(Ej \mid X_j) * P(Xj \mid X_{j-1})$$

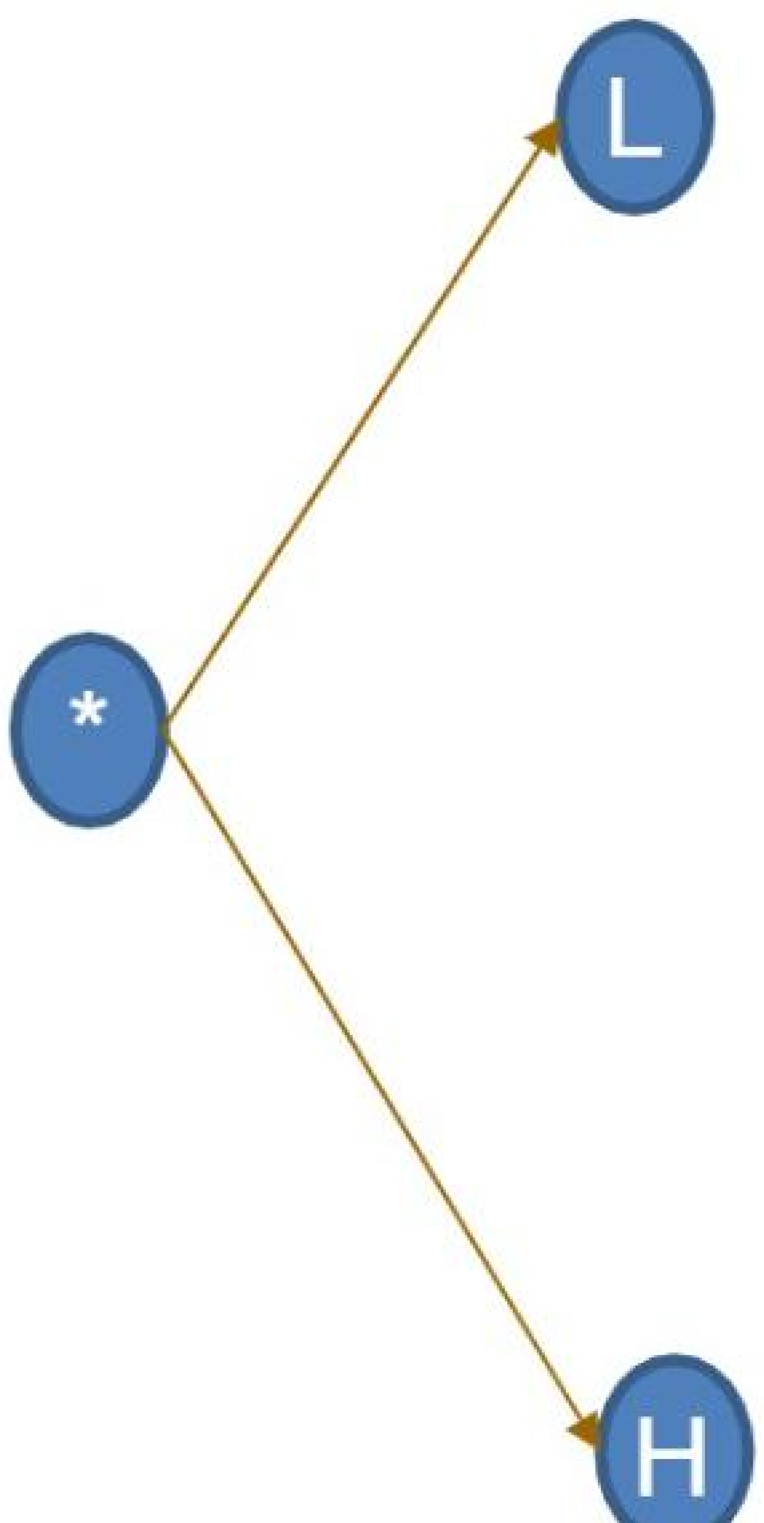
Hidden Morkov Model

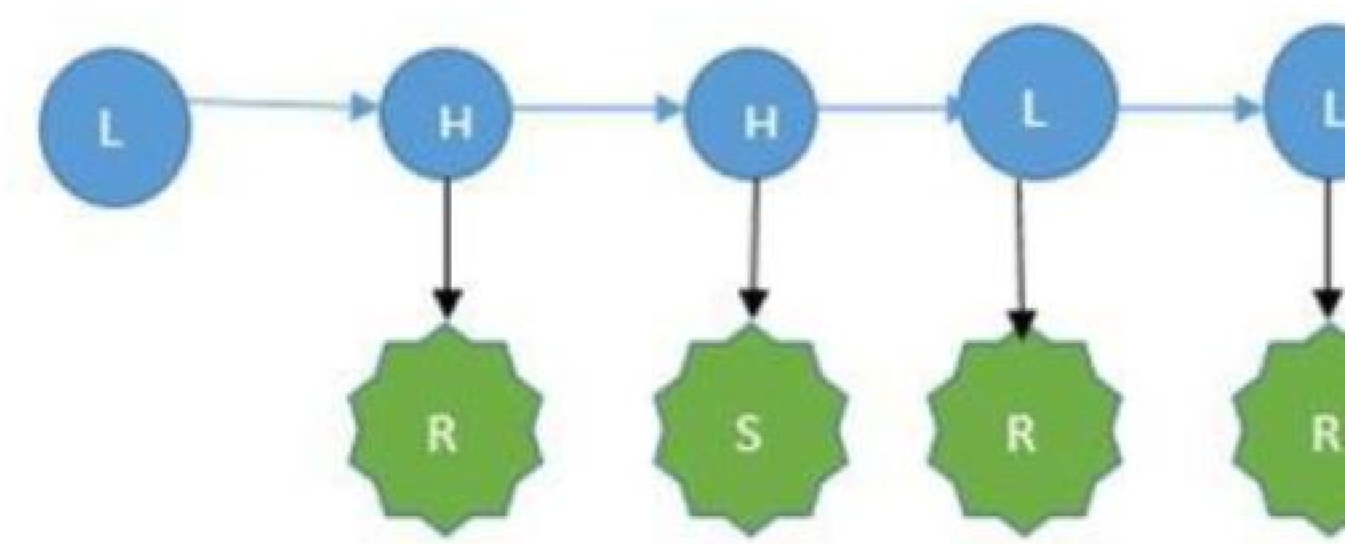
Forward Propagation Algorithm

Find the probability of occurrence of this Pressure sequence observation: **S-S-R**

<u>Initialization Phase:</u>

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

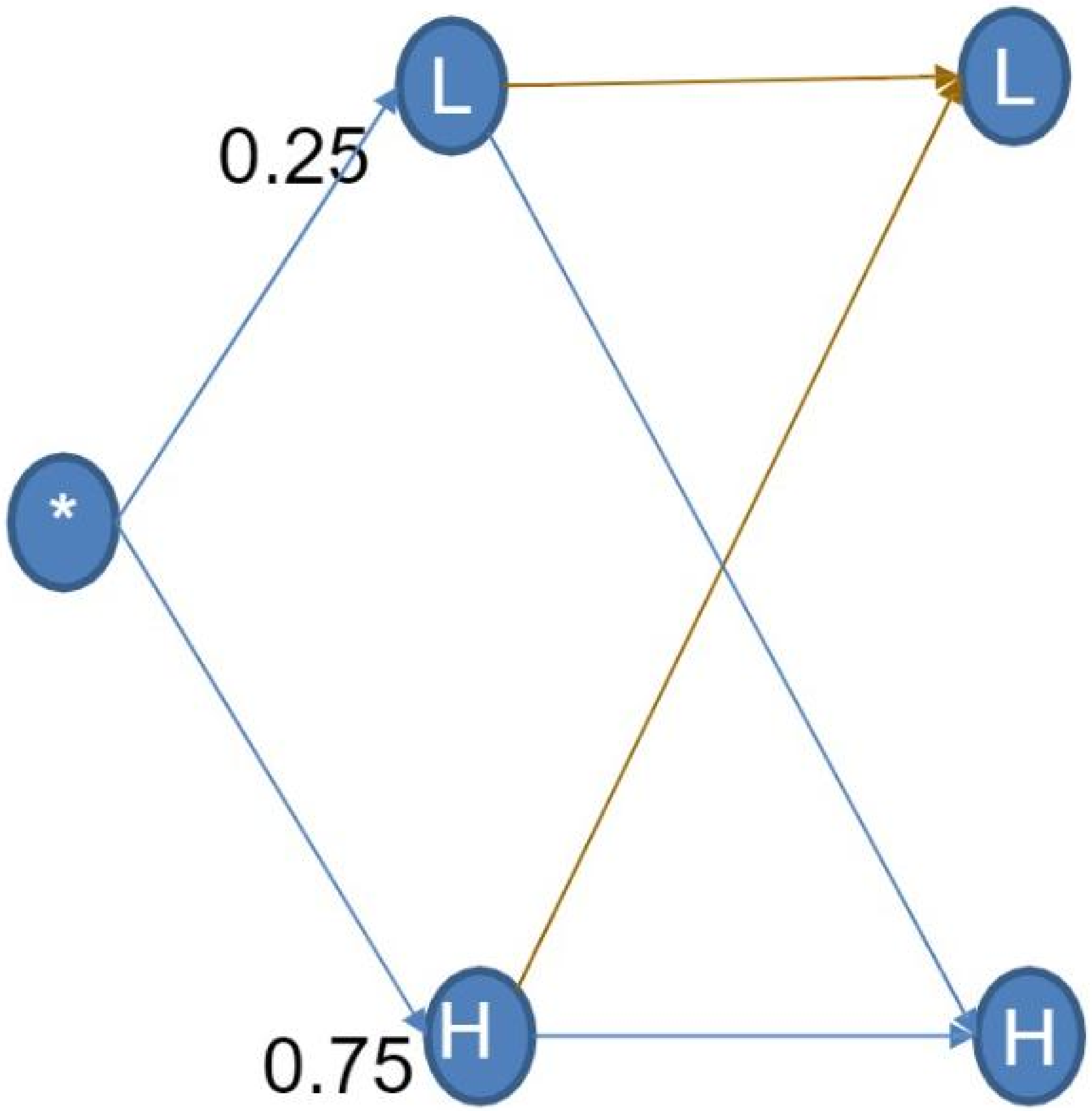
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

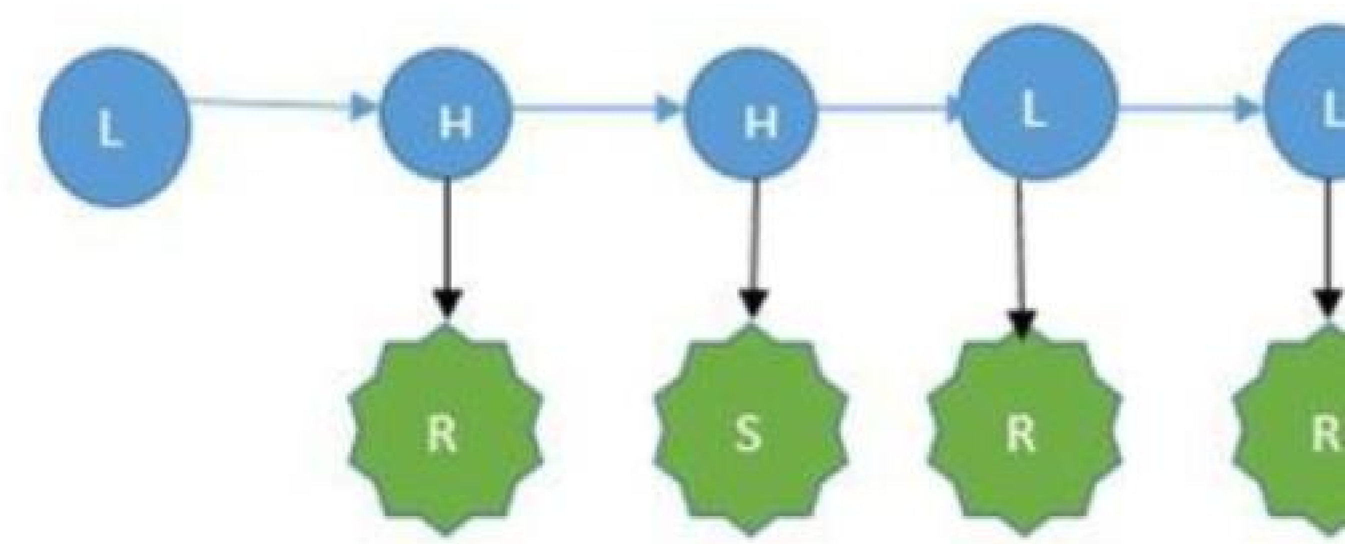
Hidden Morkov Model

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**





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

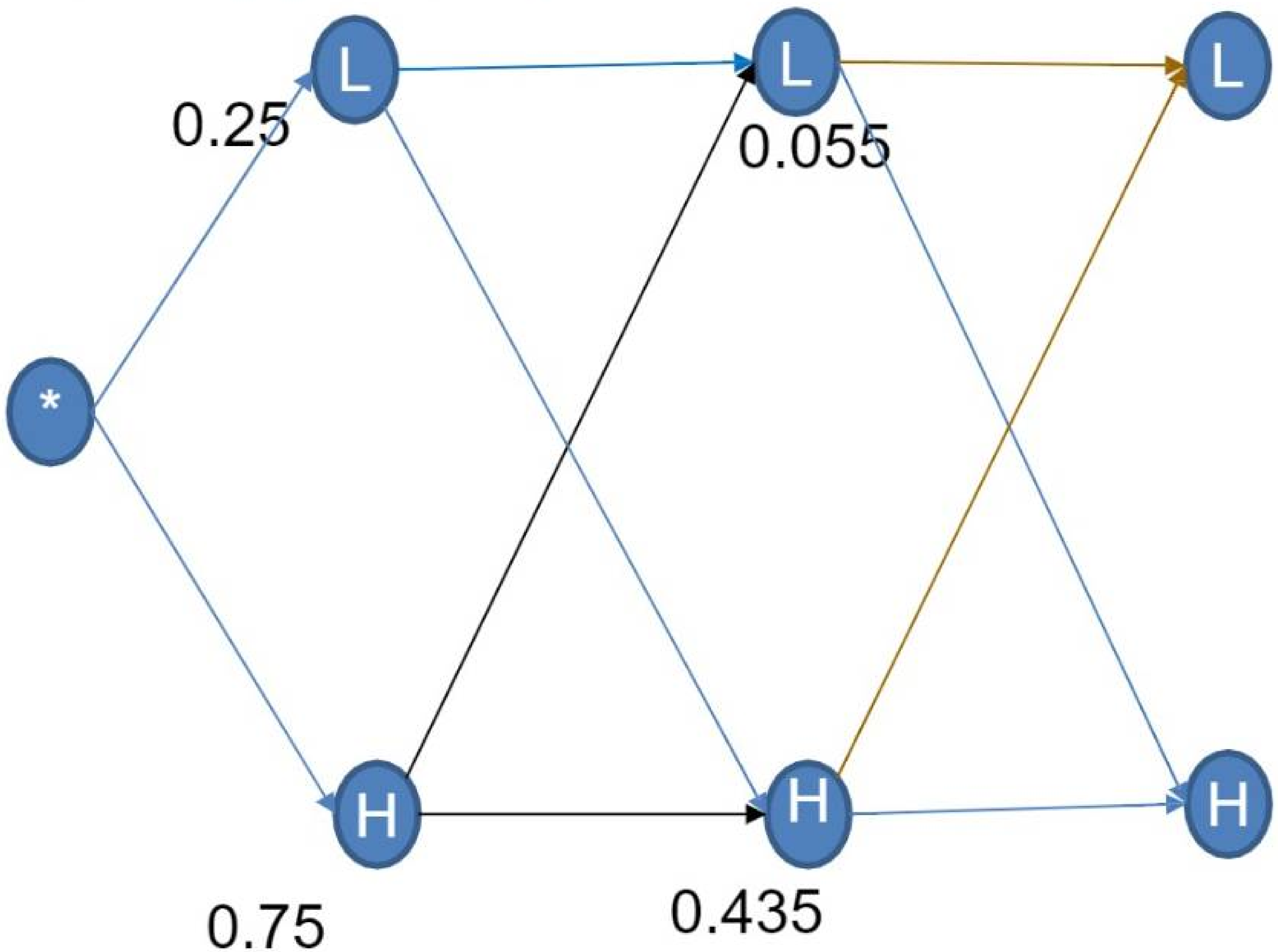
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

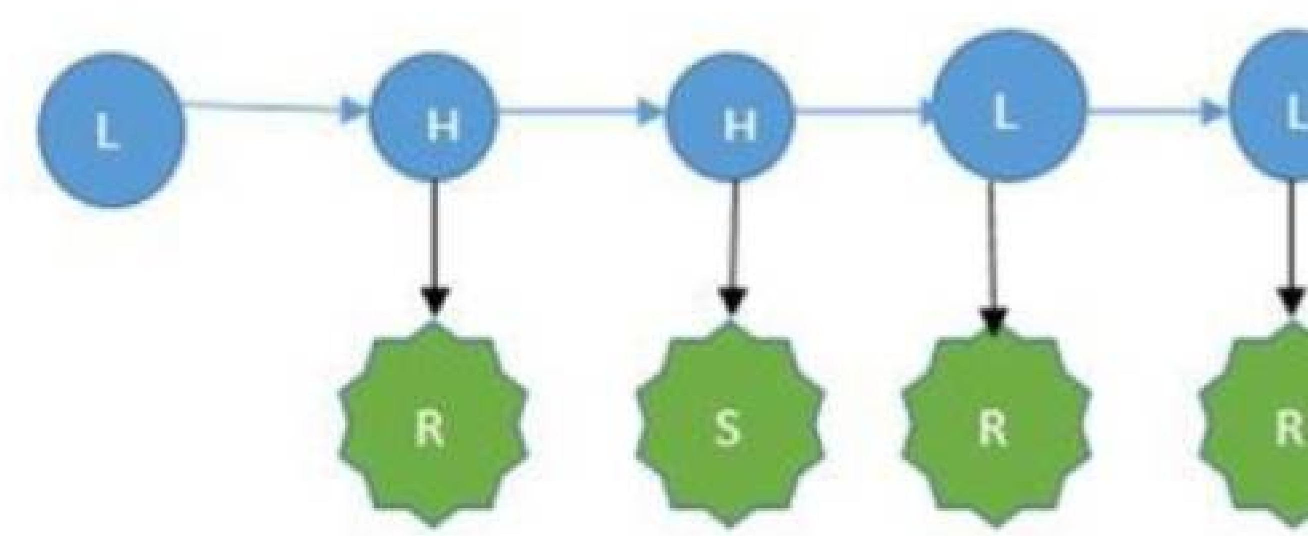
P(L)*P(L|L)*P(R|L) = 0.055*0.5*0.8 = **0.022**

P(H)*P(L|H)*P(R|L) = 0.435*0.2*0.8 = **0.0696**



P(L)*P(H|L)*P(R|H) = 0.055*0.5*0.4 = **0.011**

P(H)*P(H|H)*P(R|H) = 0.435*0.8*0.4 = **0.1392**



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

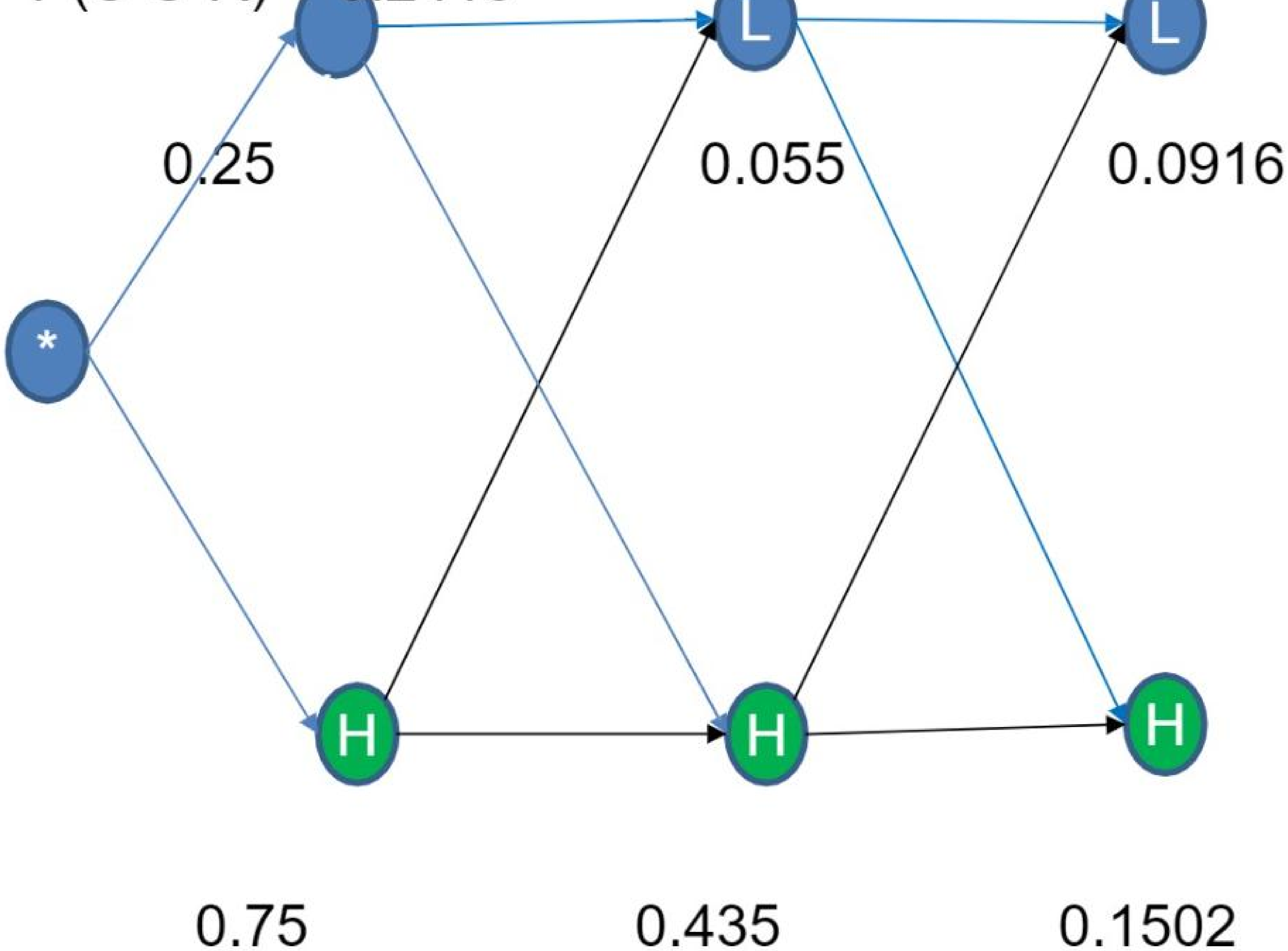
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

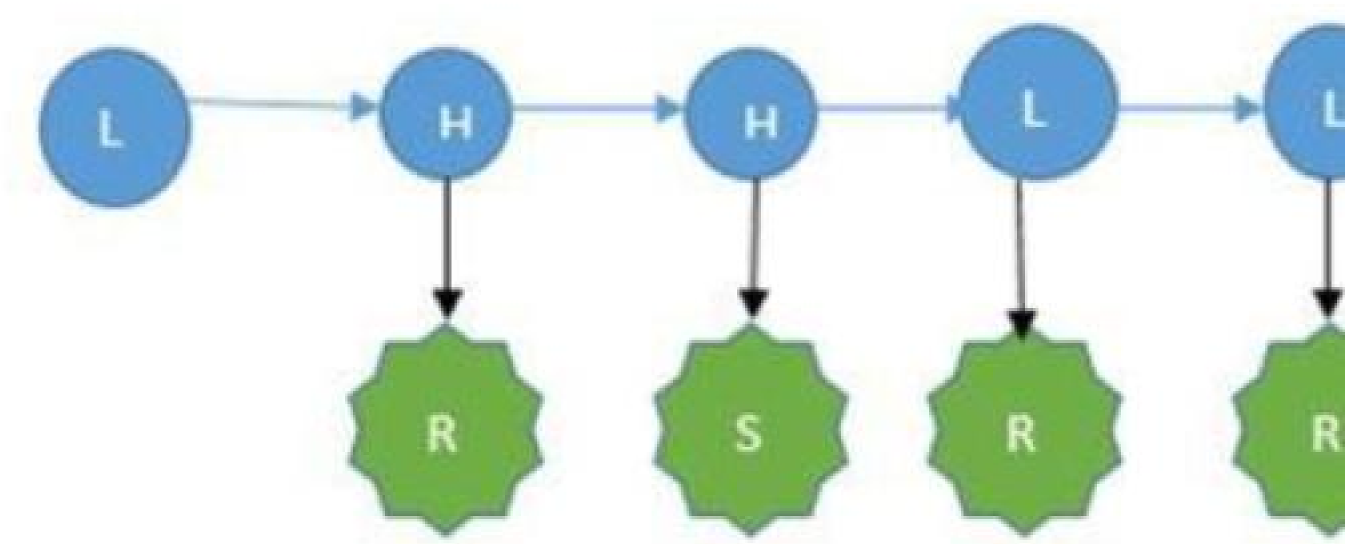
Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

<u>Termination Phase:</u>

P(S-S-R) = 0.2418





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

Filtering : P(SecondUrnIsSelected $_3$ | Red-Blue-Blue-Yellow)

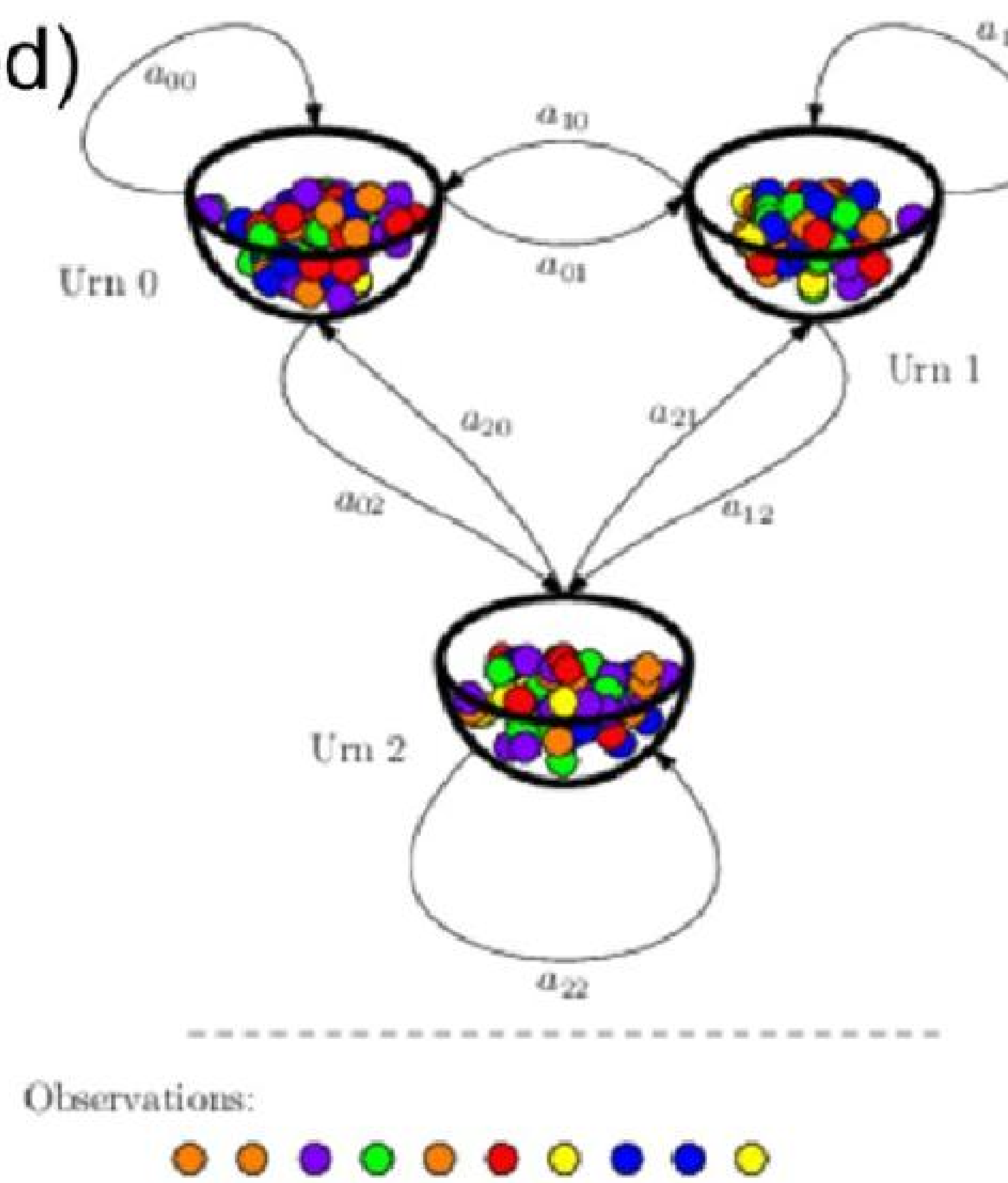
$P(X_t \mid E_{1\ldots t})$

Prediction: P(FirstUrnWillbeSelected $_3$ | Red-Yellow)

$P(X_{t+k} \mid E_{1\ldots t})$

Smoothing: P(ThirdUrnWasSelected $_2$ | Red-Yellow-Red-Red)

$P(X_{k,\ o>k>t} \mid E_{1\ldots t})$



Most Likely Explanation (or) Viterbi Algorithm

P(Urn1-Urn2-Urn1 | Red-Yellow-Yellow)

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

Hidden Morkov Model

Inference: Type -2

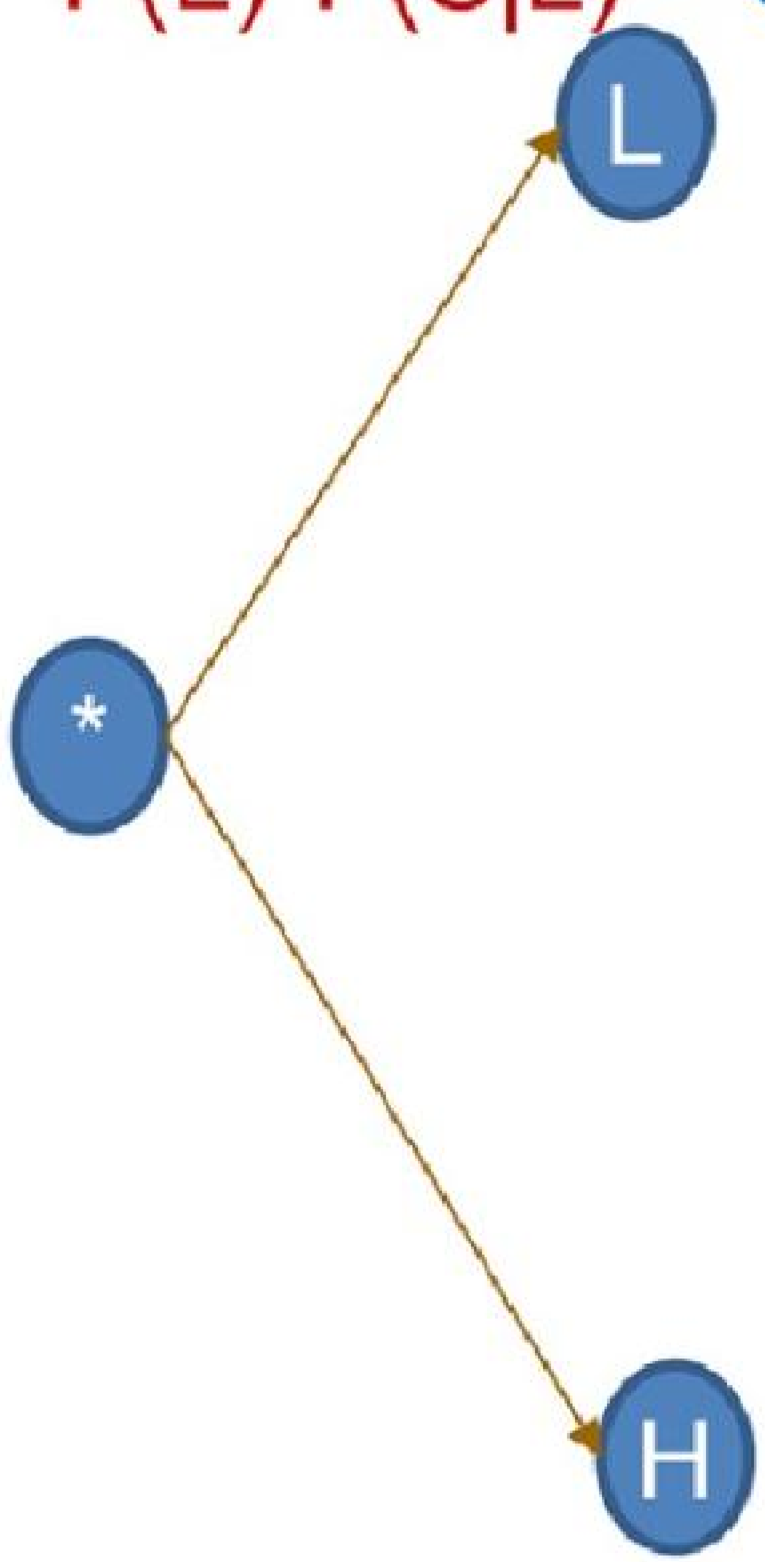
Most Likely Explanation : Veterbi Algorithm

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

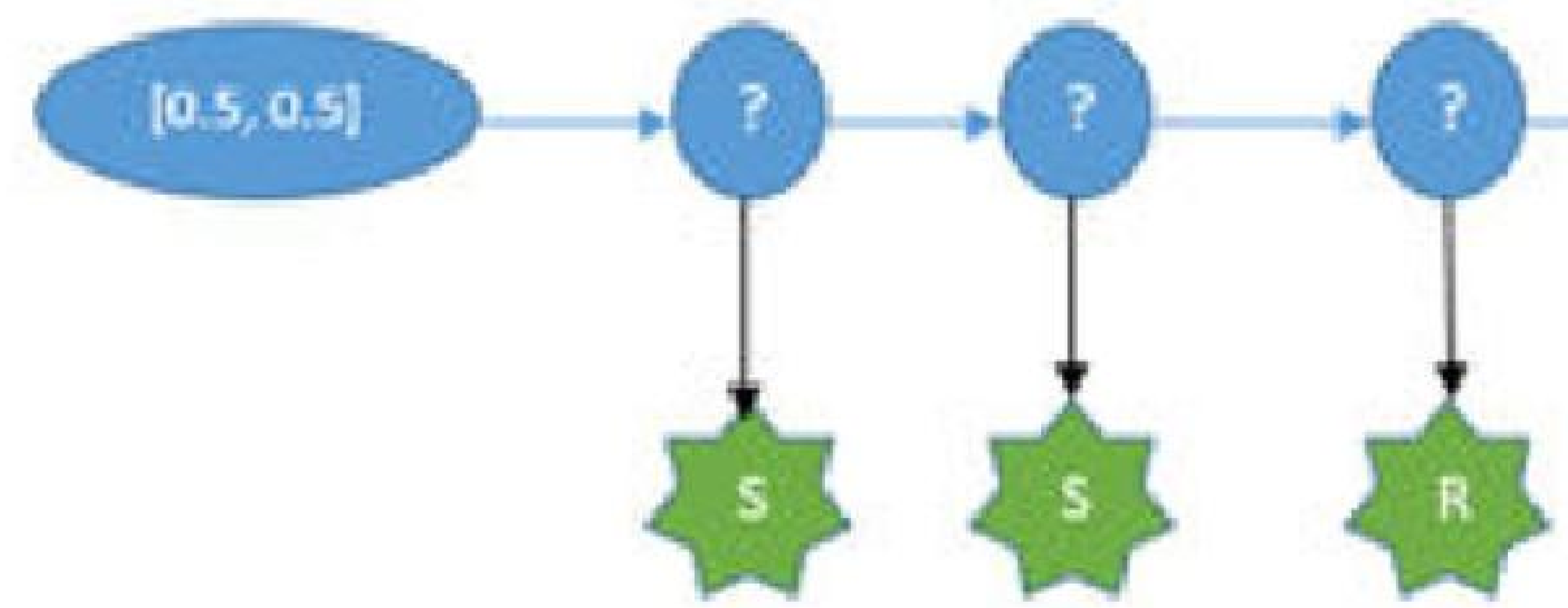
argmax $X_{1...t}$: $P(X_{1...t} \mid E_{1...t})$

MM Inf

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



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



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

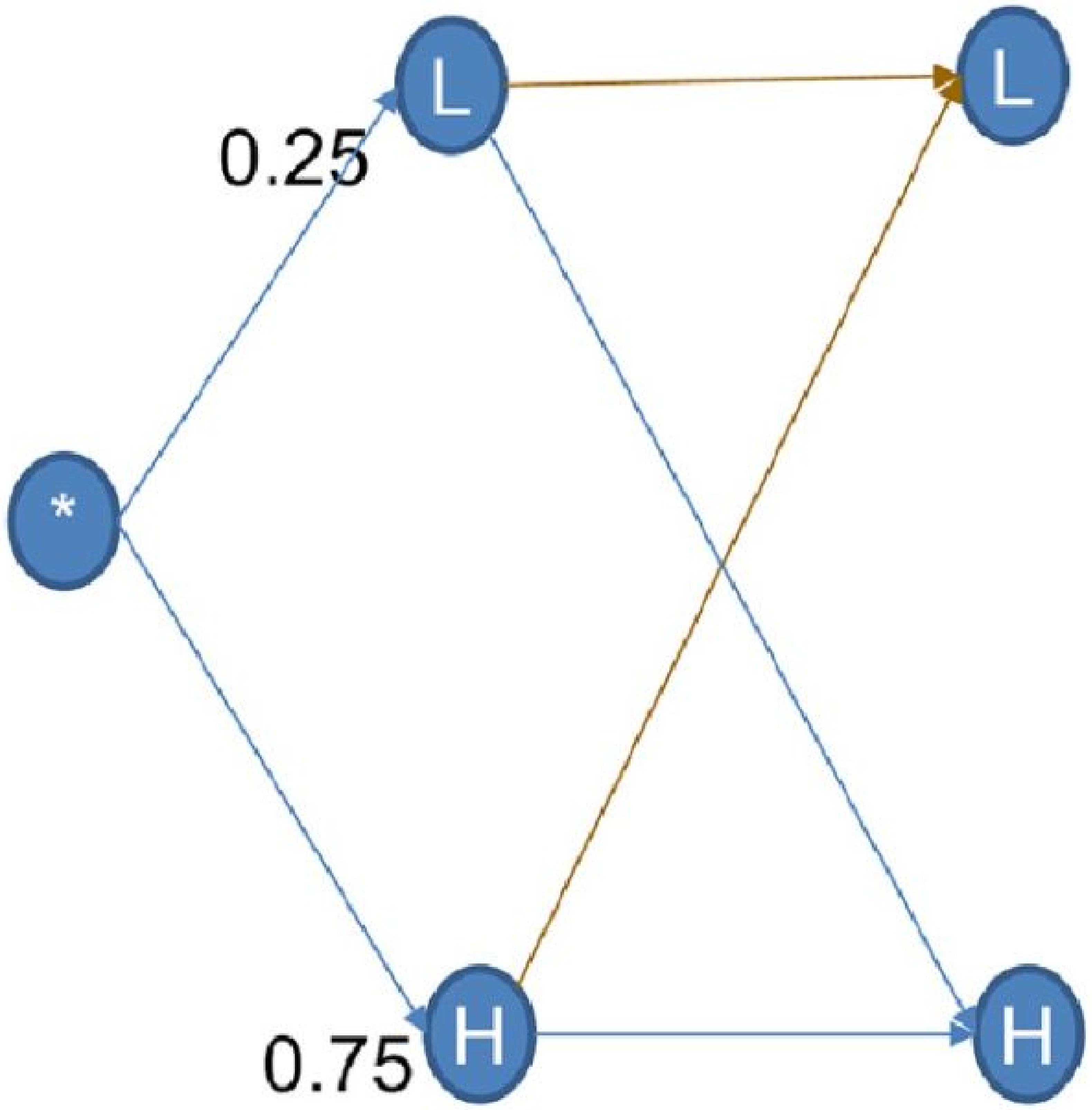
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

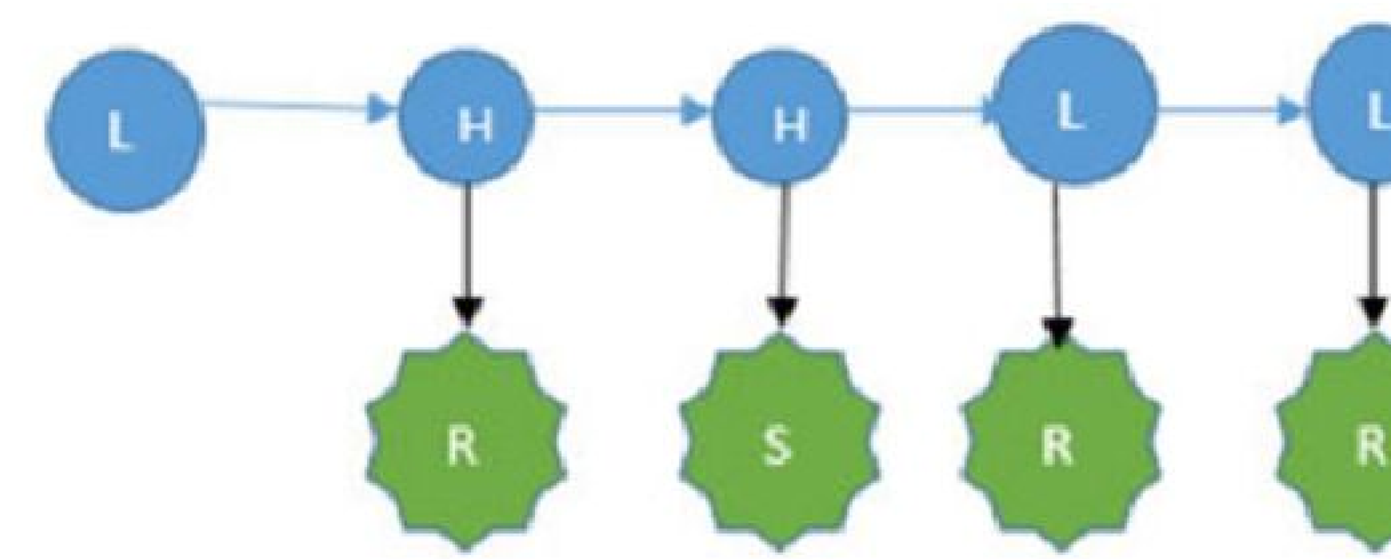
Hidden Morkov Model

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

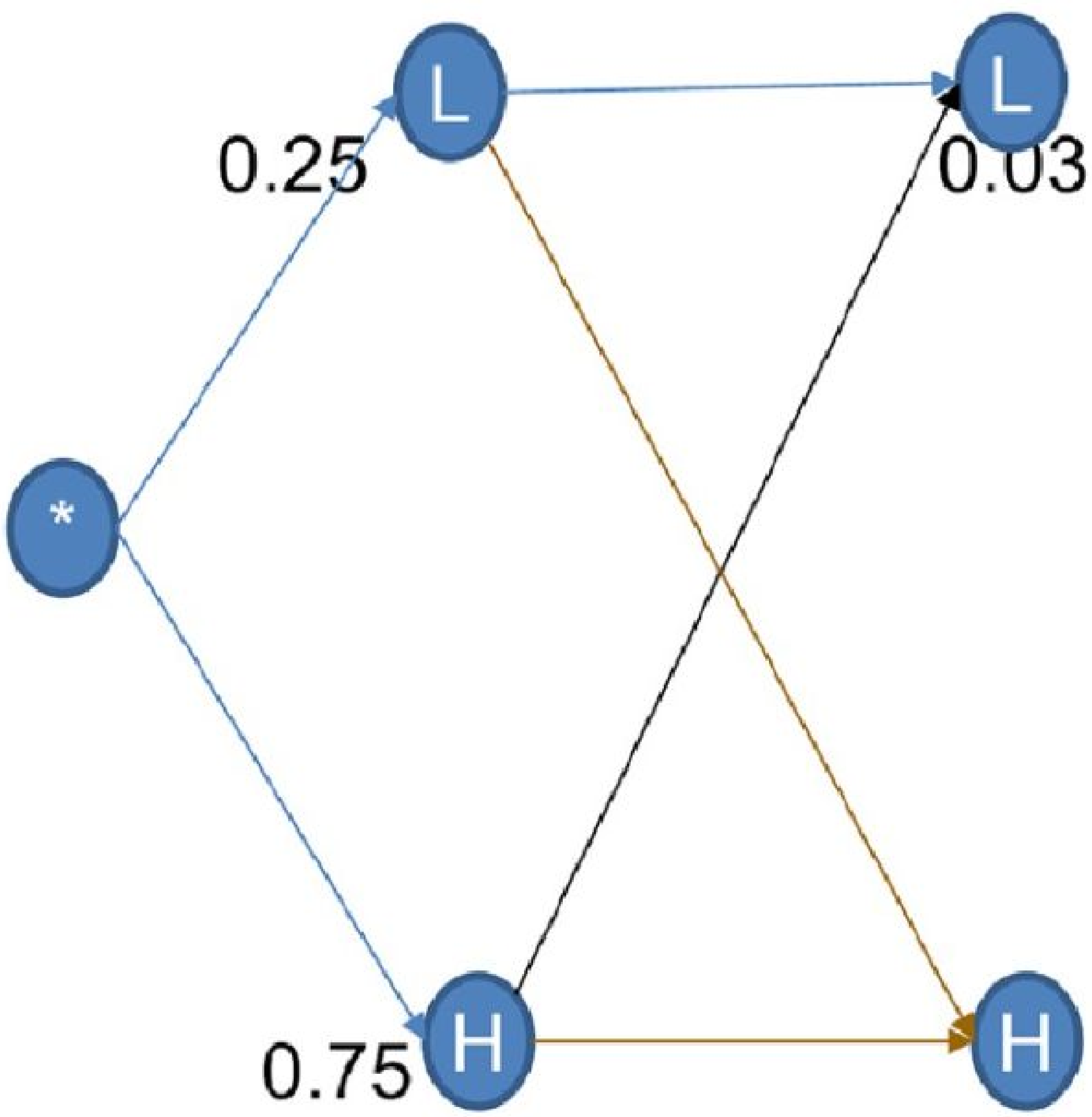
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

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

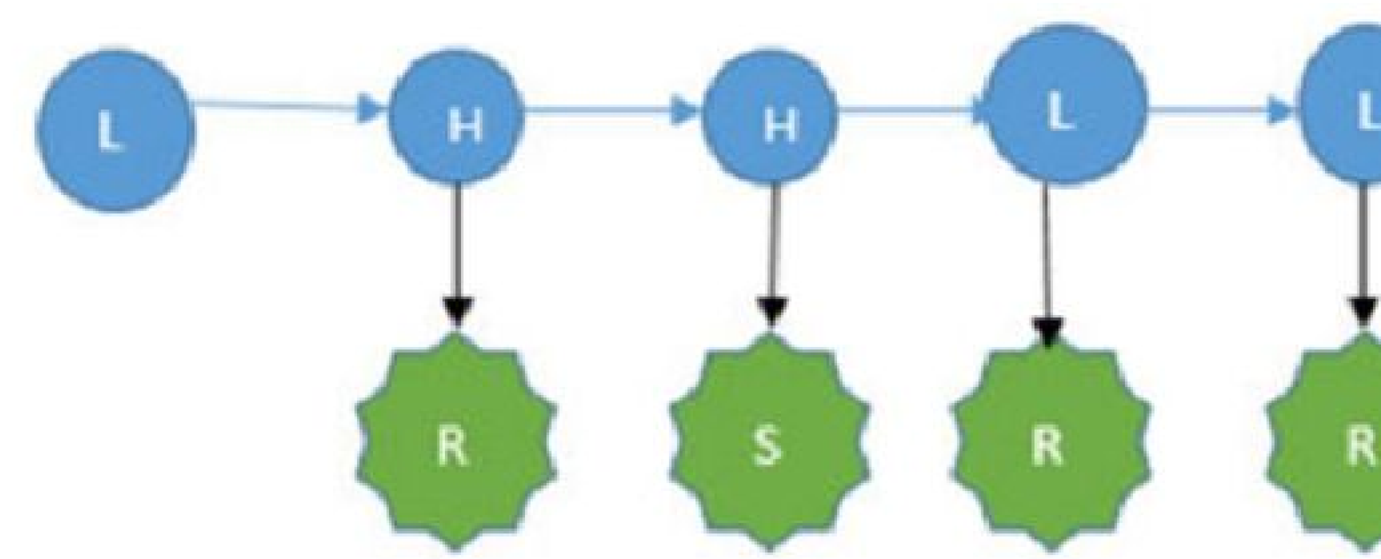
Hidden Morkov Model

Veterbi Algorithm : S-S-R



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

$$P(H)*P(H|H)*P(S|H) = 0.75*0.8*0.6 = \mathbf{0.36}$$



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

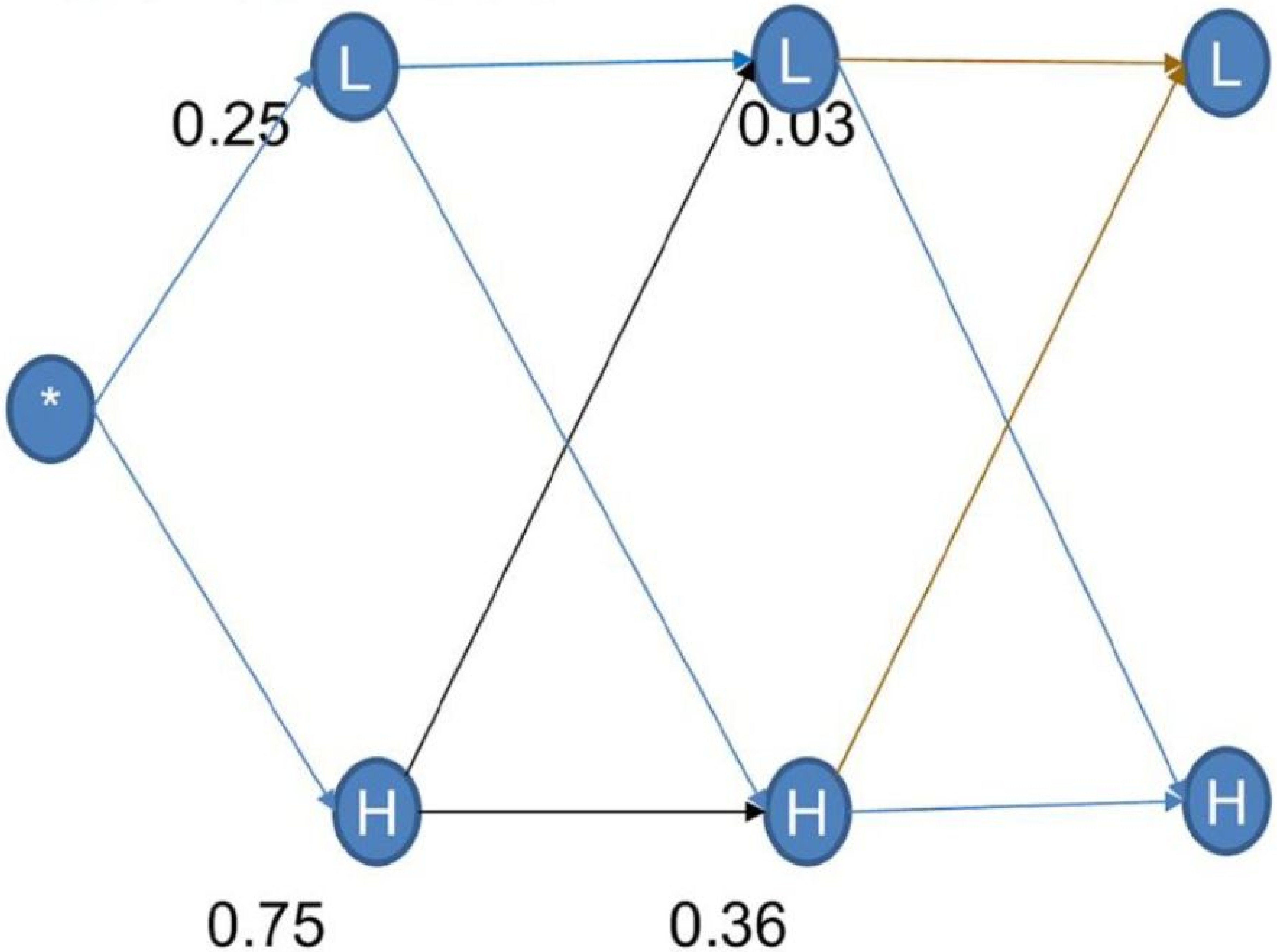
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

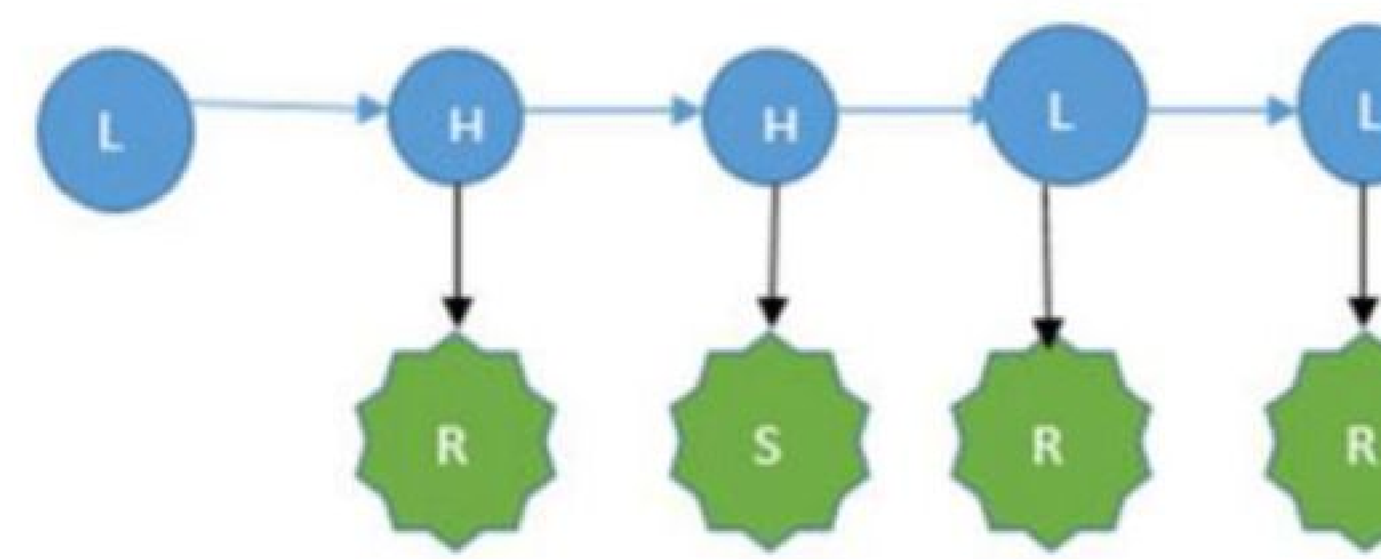
Hidden Morkov Model

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

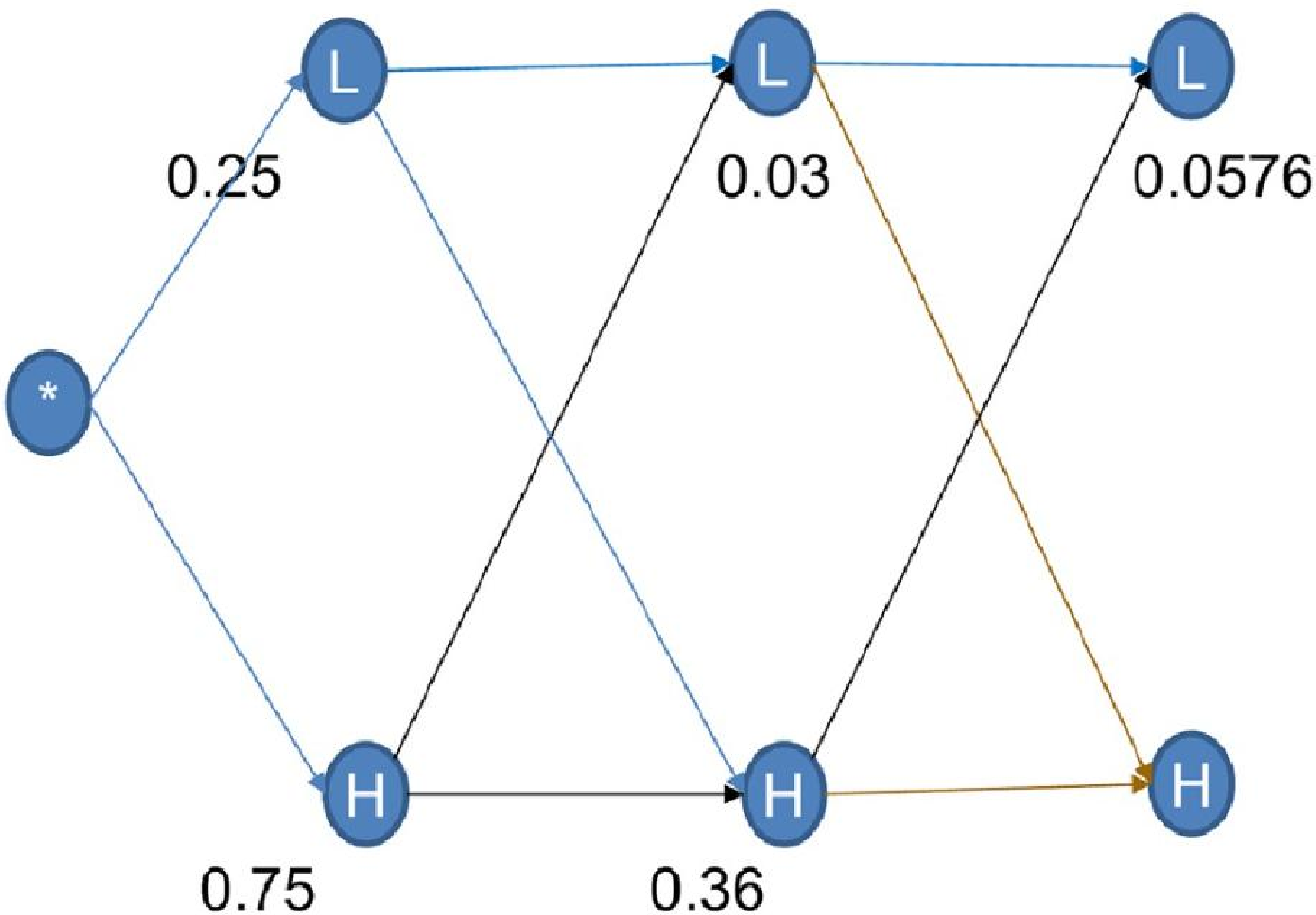
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

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

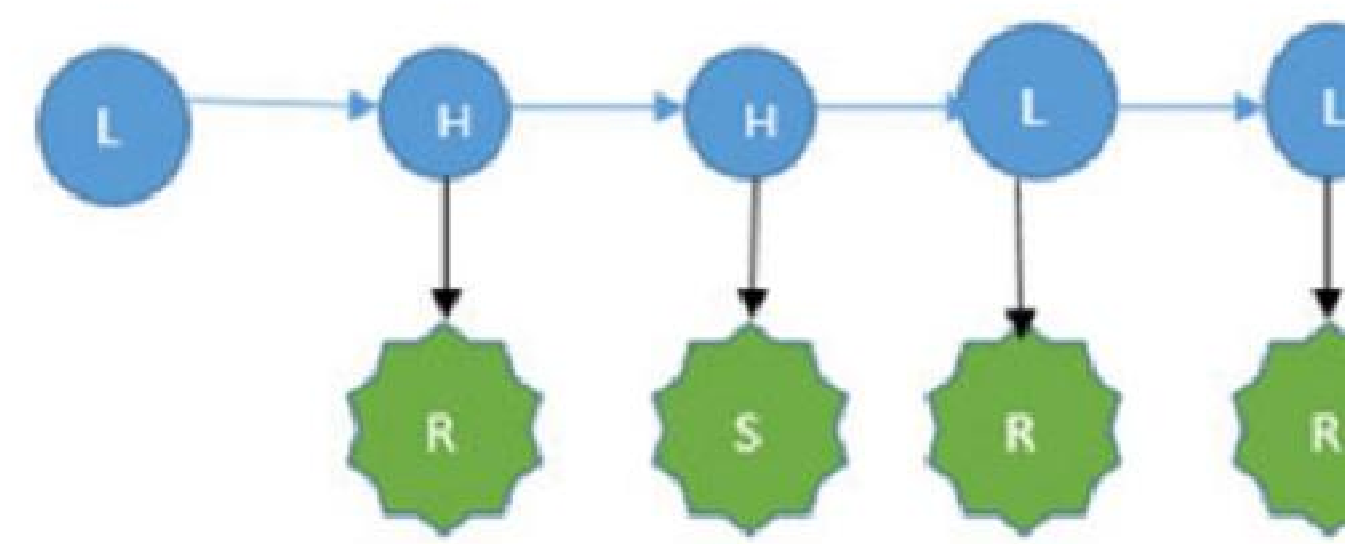
Hidden Morkov Model

Veterbi Algorithm : S-S-R



P(L)*P(H|L)*P(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

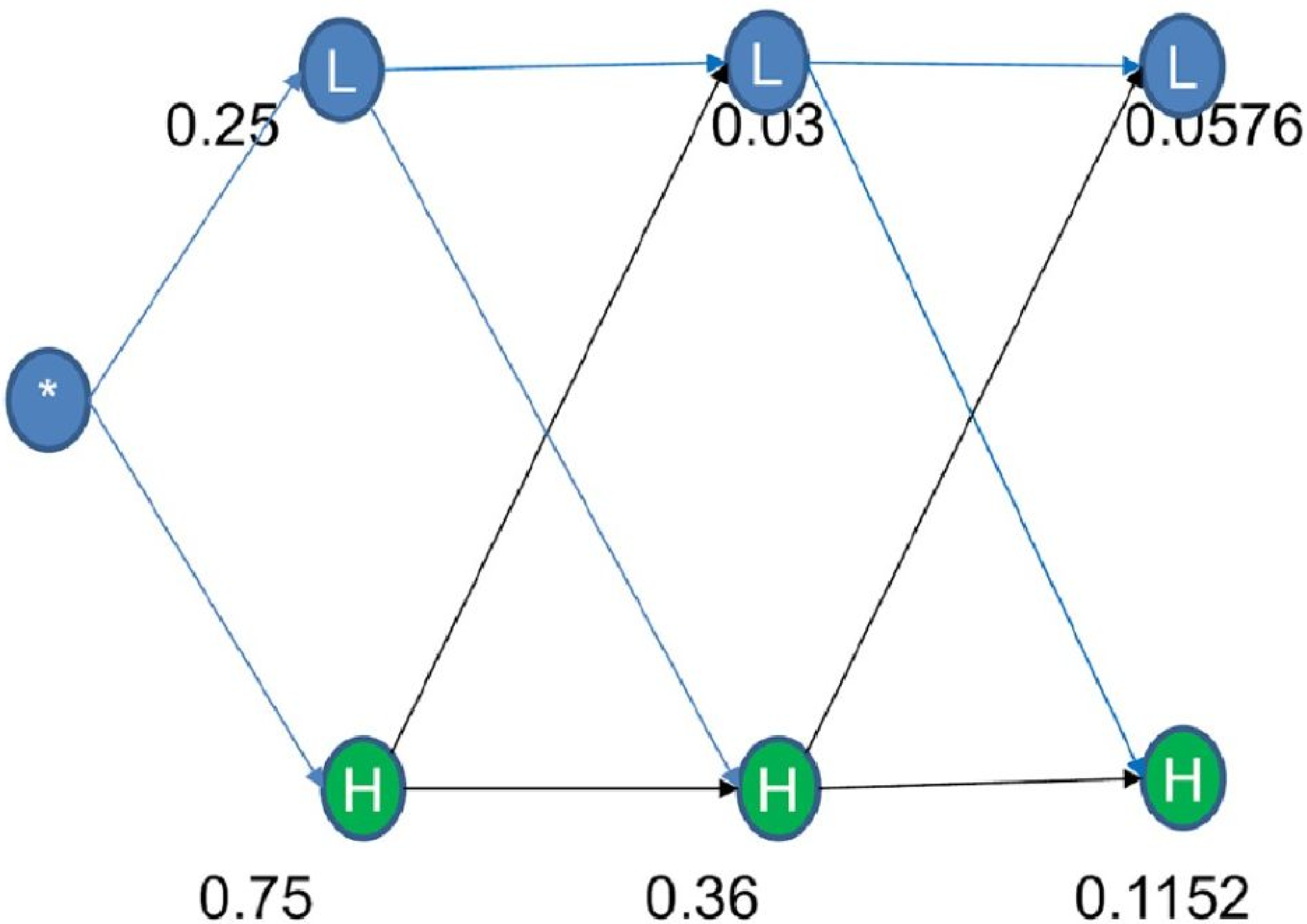
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

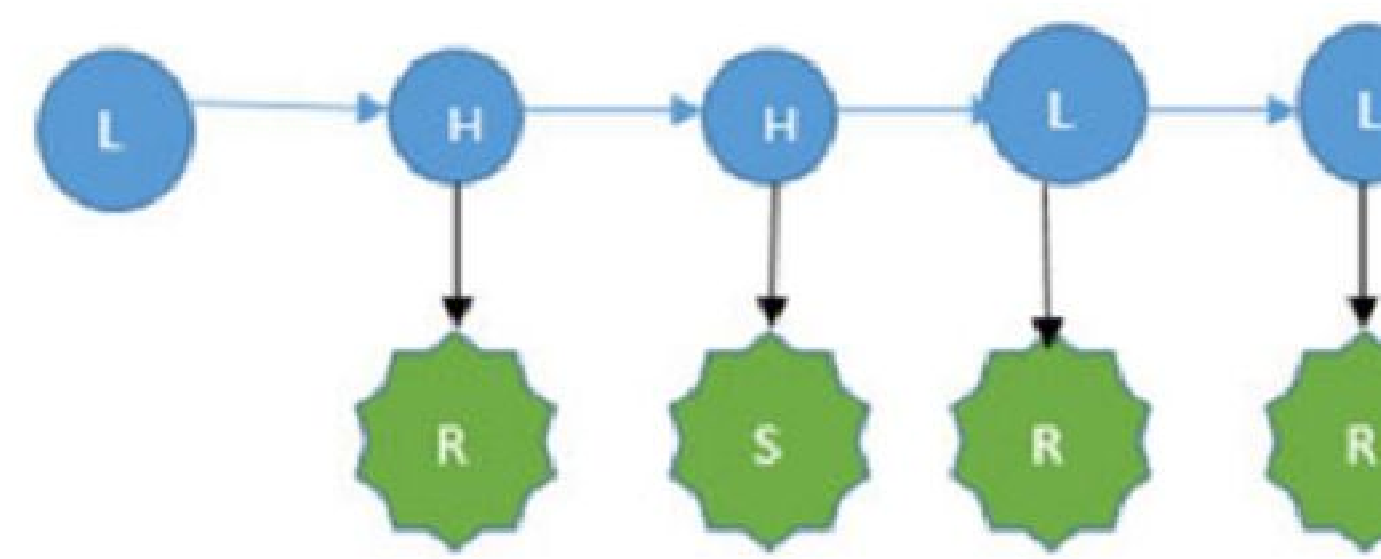
Evidence / Sensor Model/ Emission Probability Matrix

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

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	$P(E_t = \textbf{Rainy})$
0.2	0.6	$P(E_t = \textbf{Sunny})$

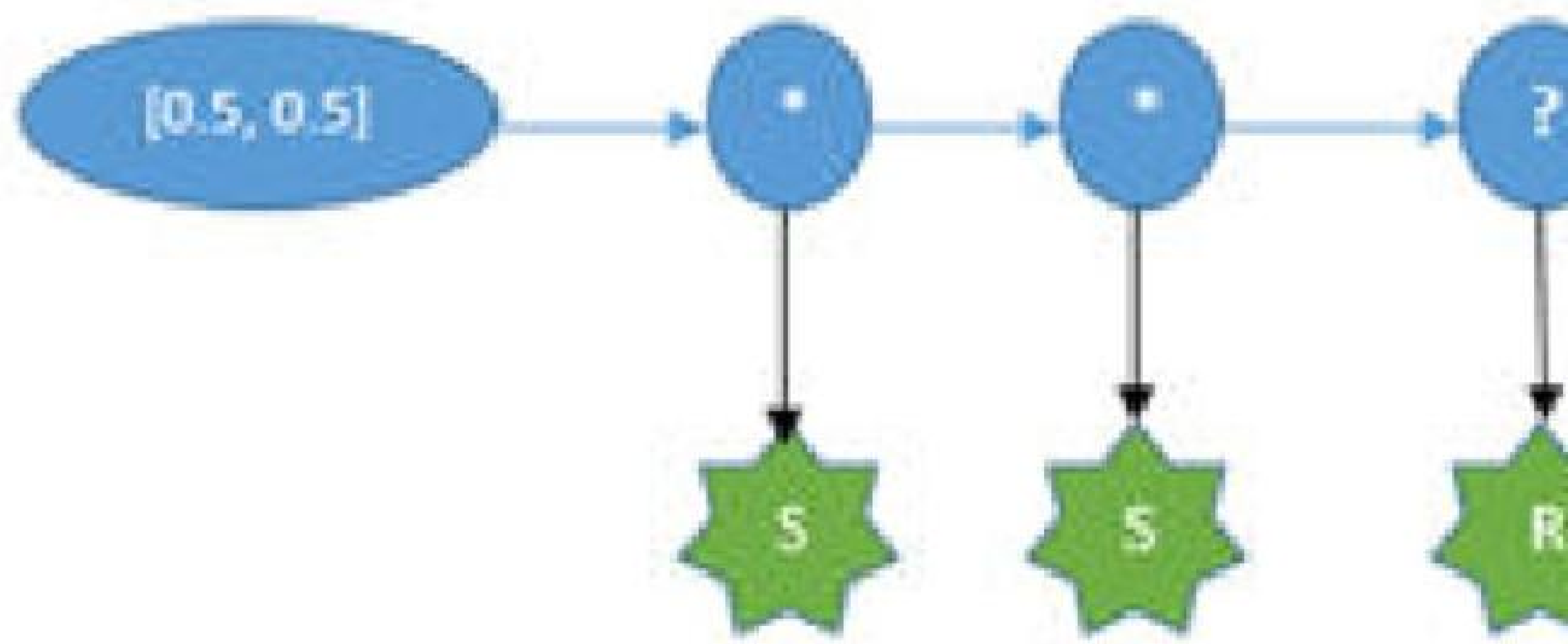
Hidden Morkov Model

Inference: Type -3

Filtering : Forward Propagation Algorithm

Find the Current Pressure if sequence of 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	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

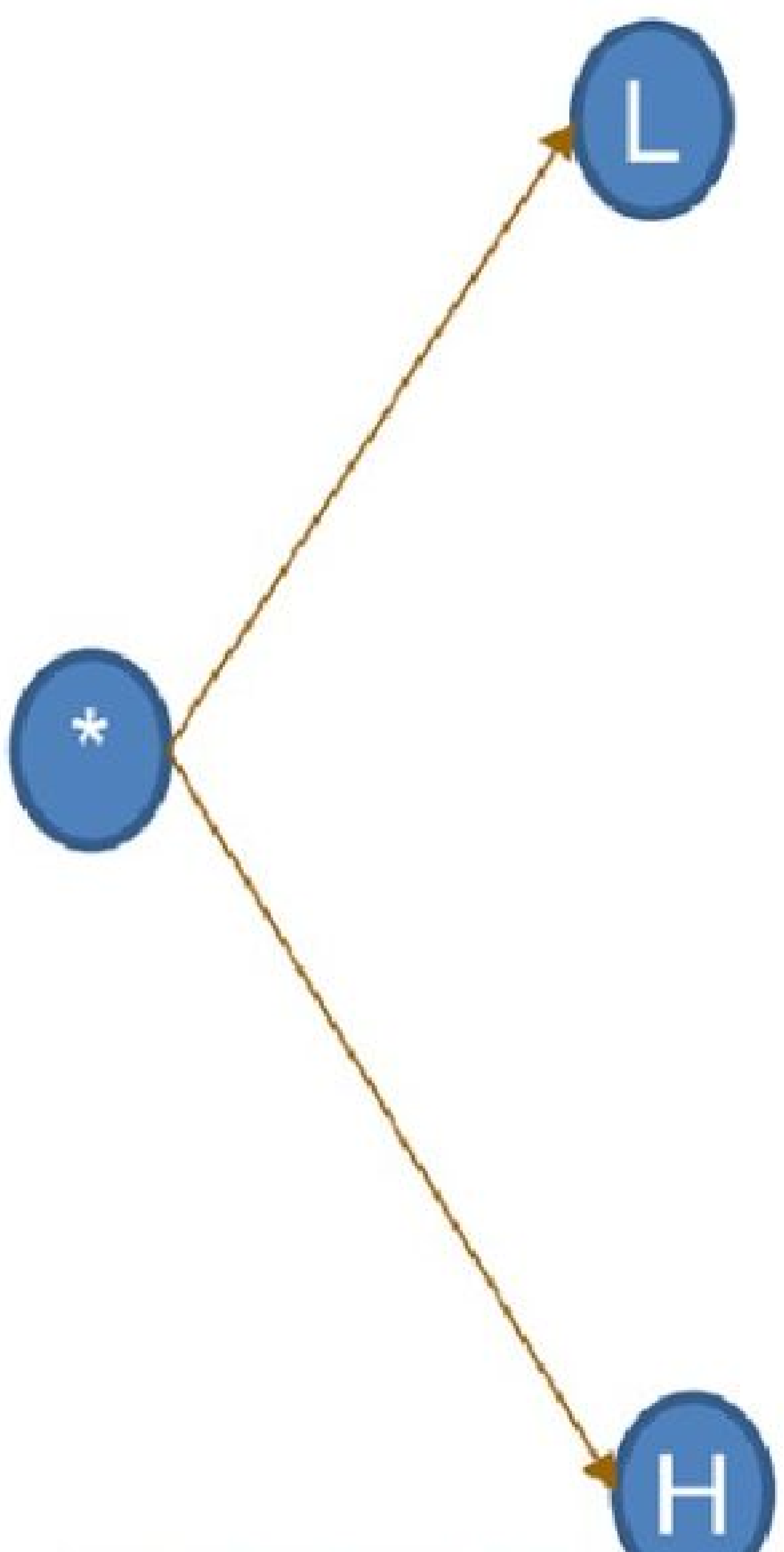
Hidden Morkov Model

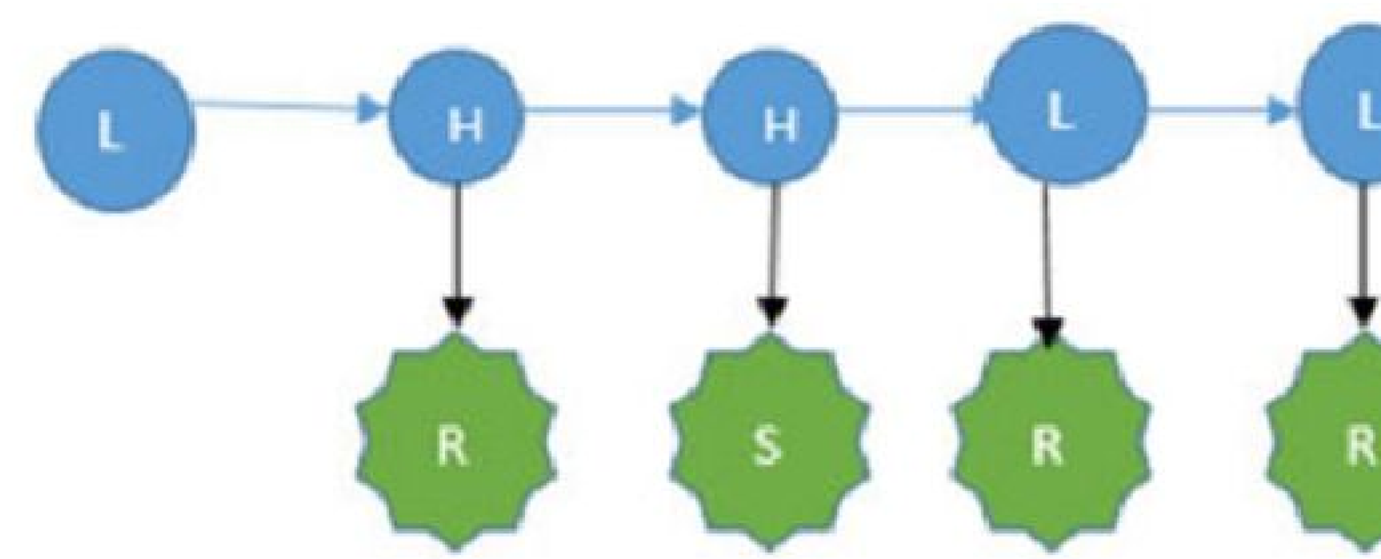
Forward Propagation Algorithm

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

Initialization Phase:

P(L)*P(S|L) = 0.5*0.2 = 0.1 → 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

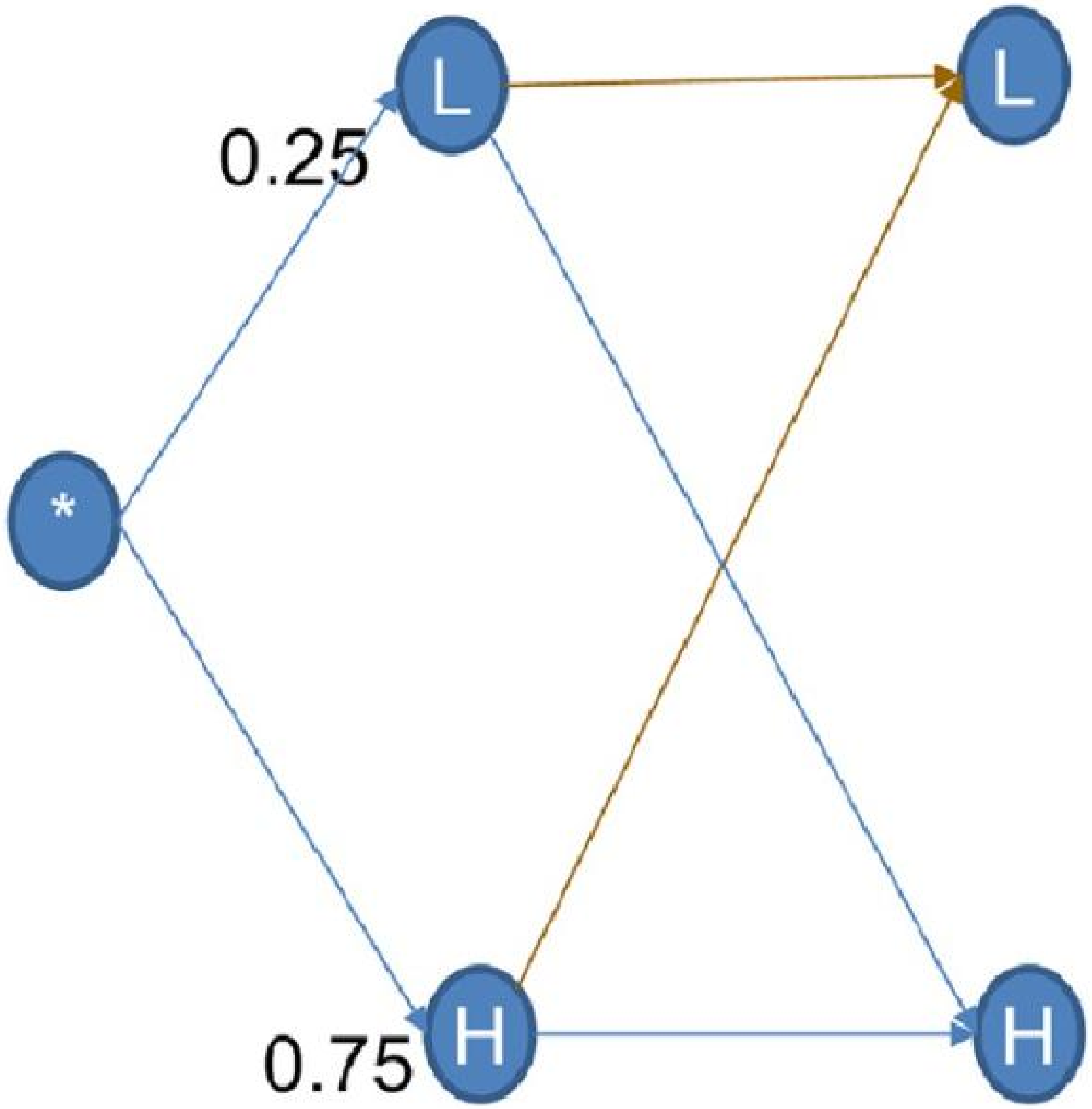
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

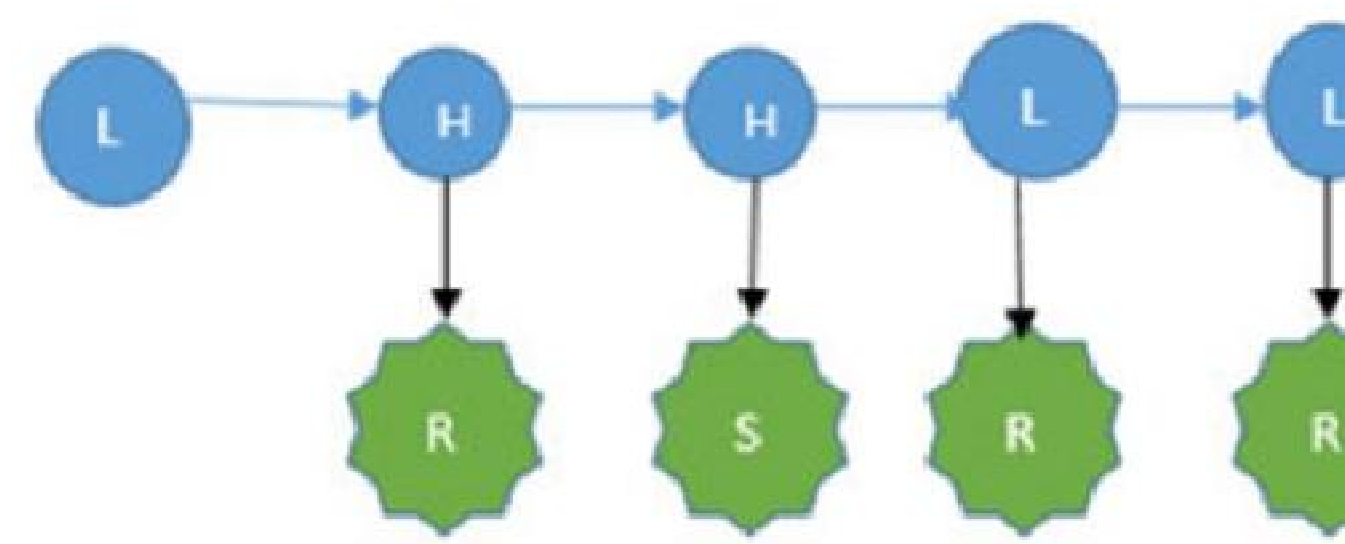
Hidden Morkov Model

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	$P(U_t = HP)$

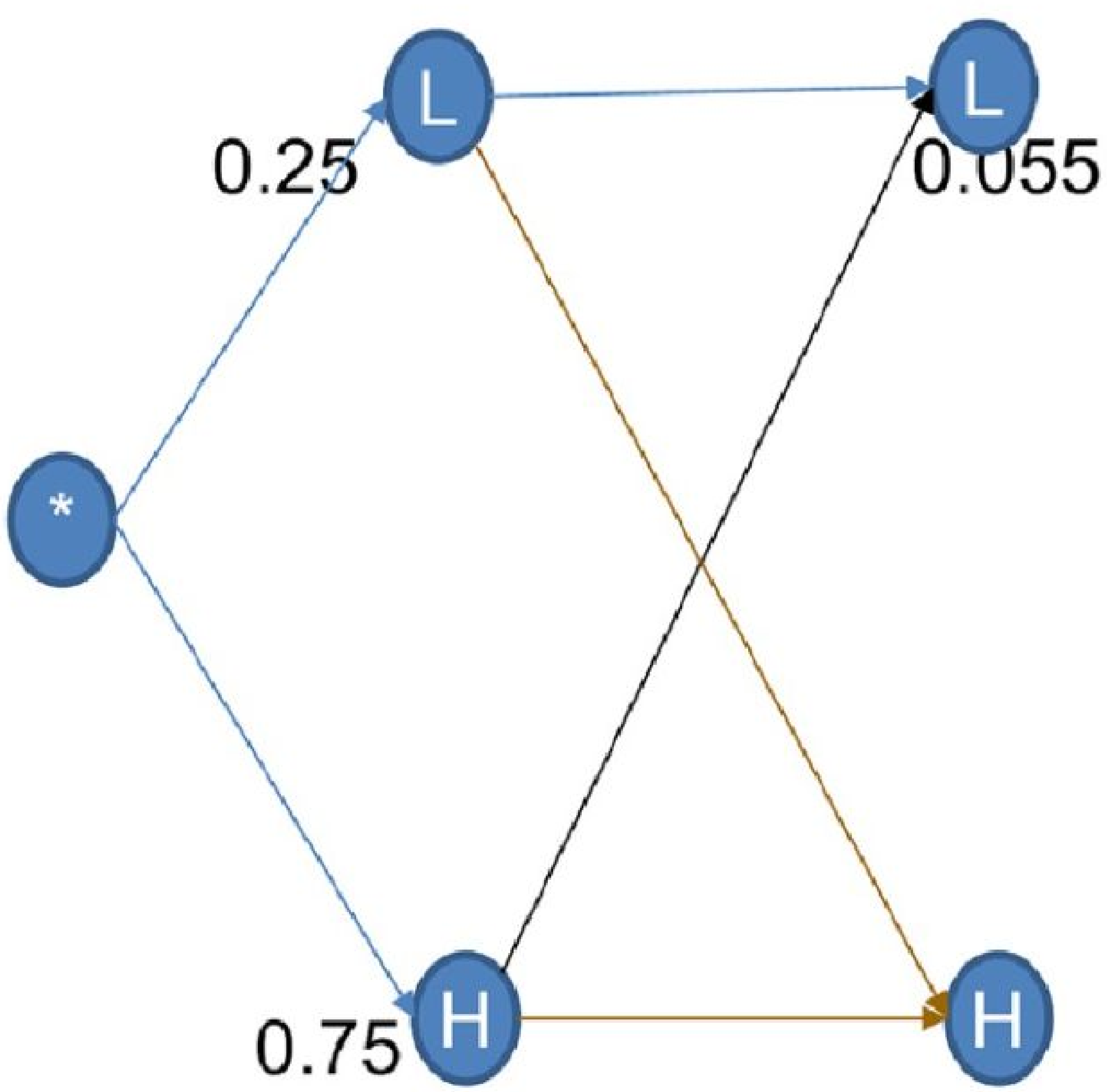
Evidence / Sensor Model/ Emission Probability Matrix

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

Recursion Phase:

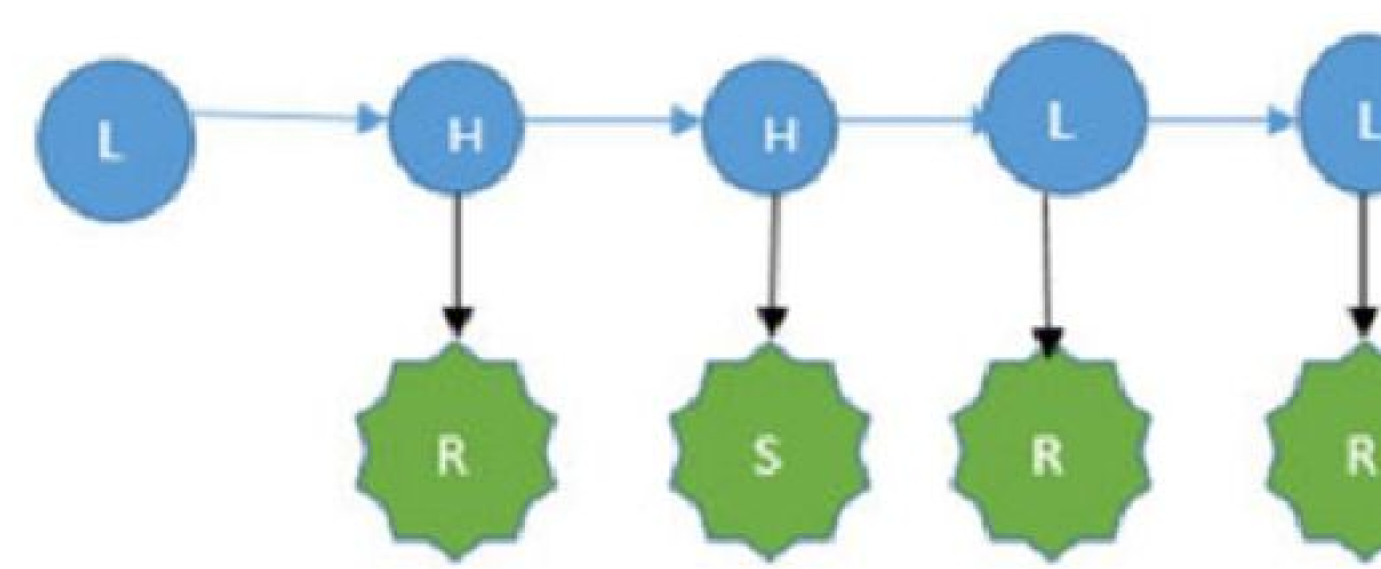
Hidden Morkov Model

Forward Propagation Algorithm : S-S-R



$P(L)*P(H|L)*P(S|H) = 0.25*0.5*0.6 = \mathbf{0.075}$

$P(H)*P(H|H)*P(S|H) = 0.75*0.8*0.6 = \mathbf{0.36}$



Transition Model / Probability Matrix

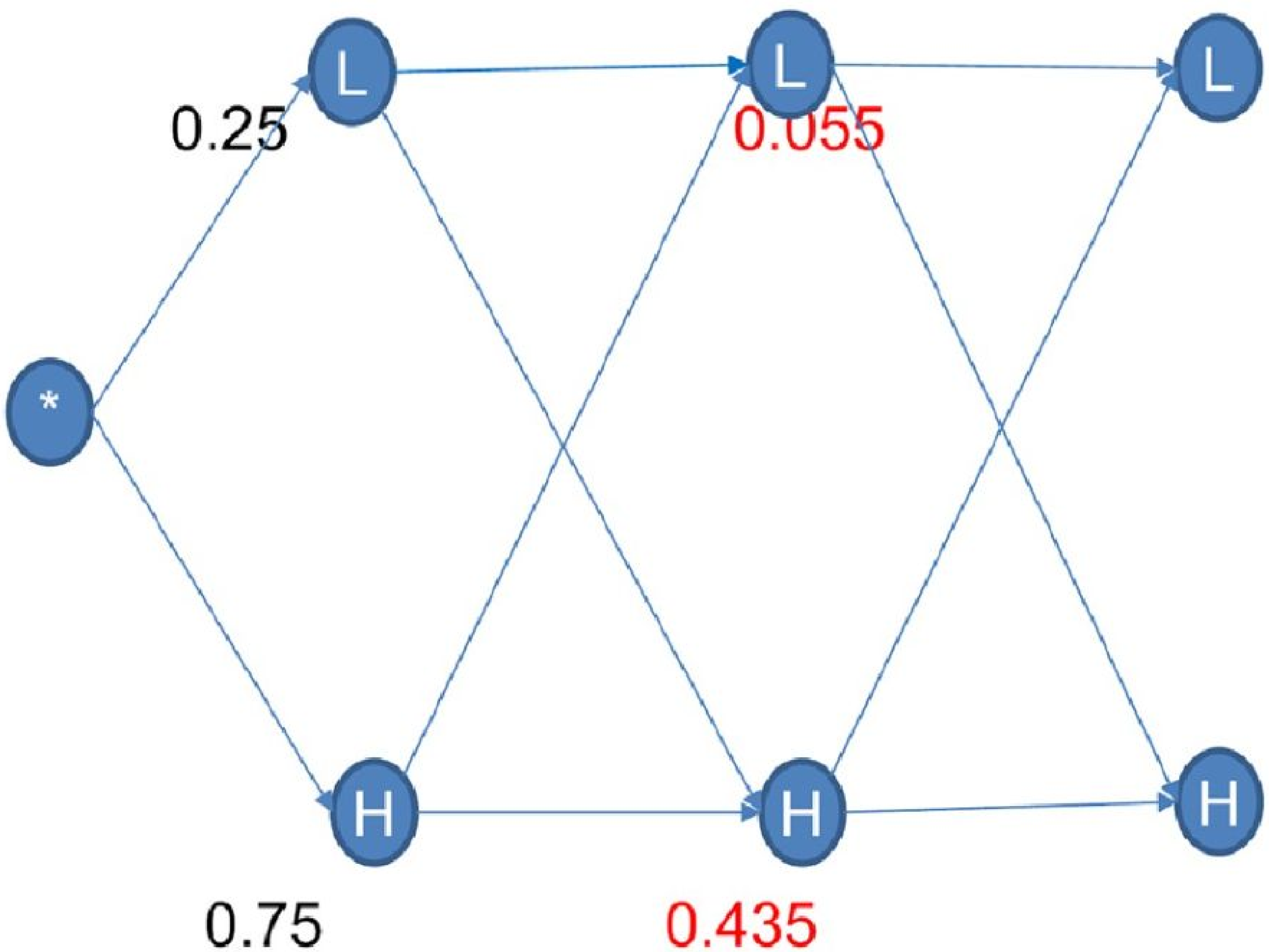
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

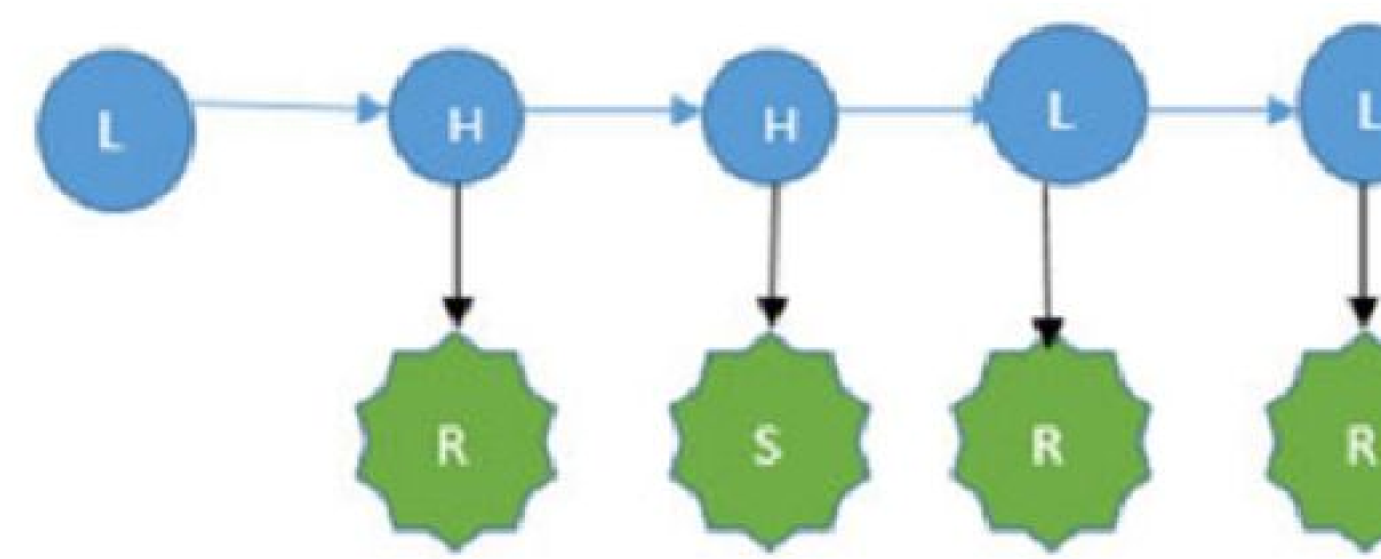
Evidence / Sensor Model/ Emission Probability Matrix

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

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

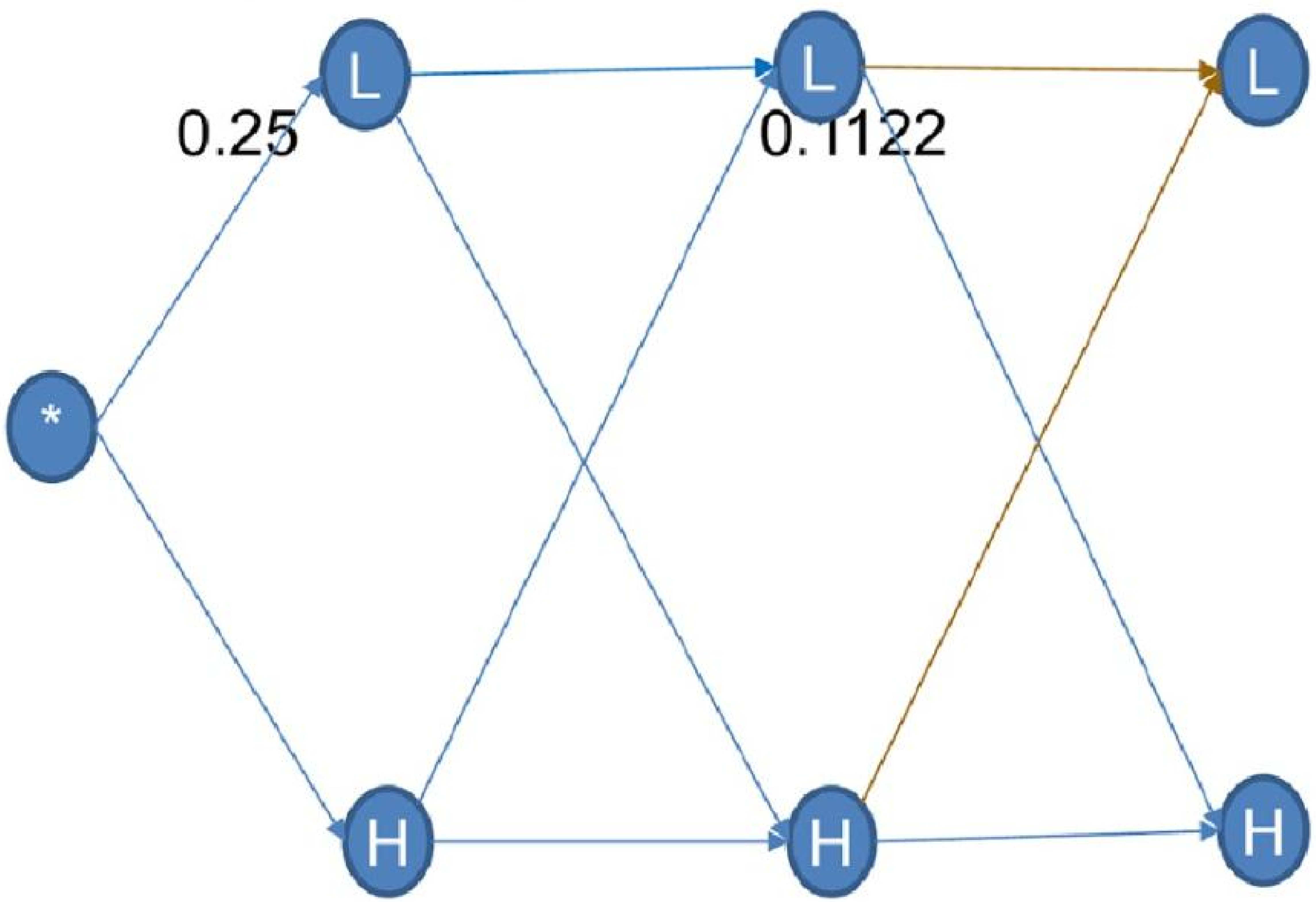
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

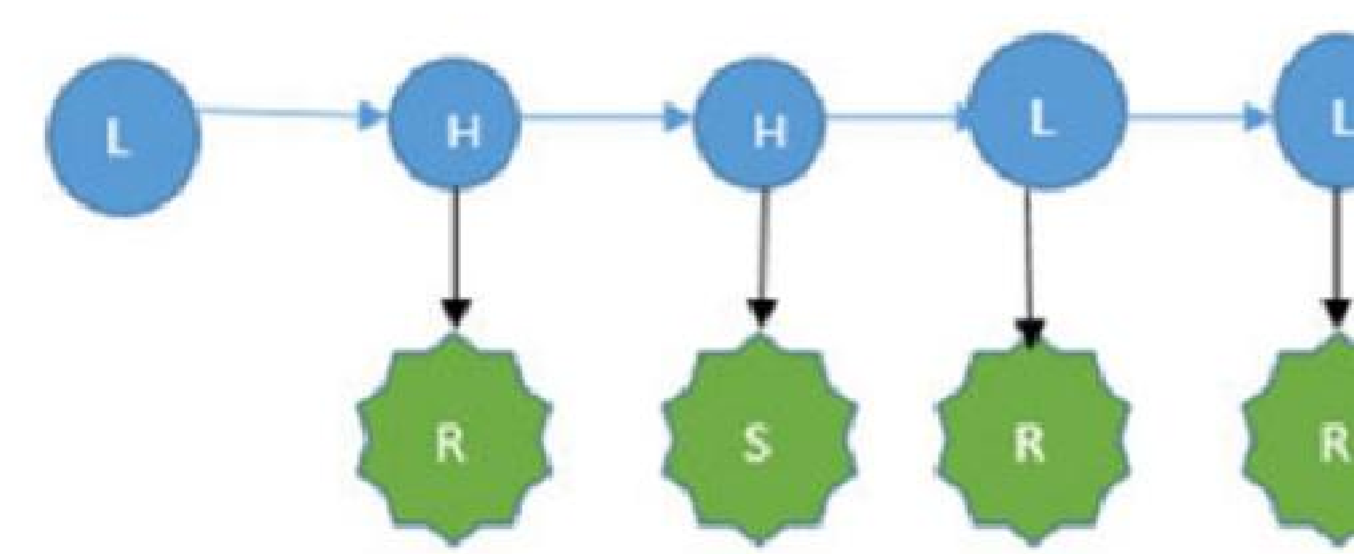
Hidden Morkov Model

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

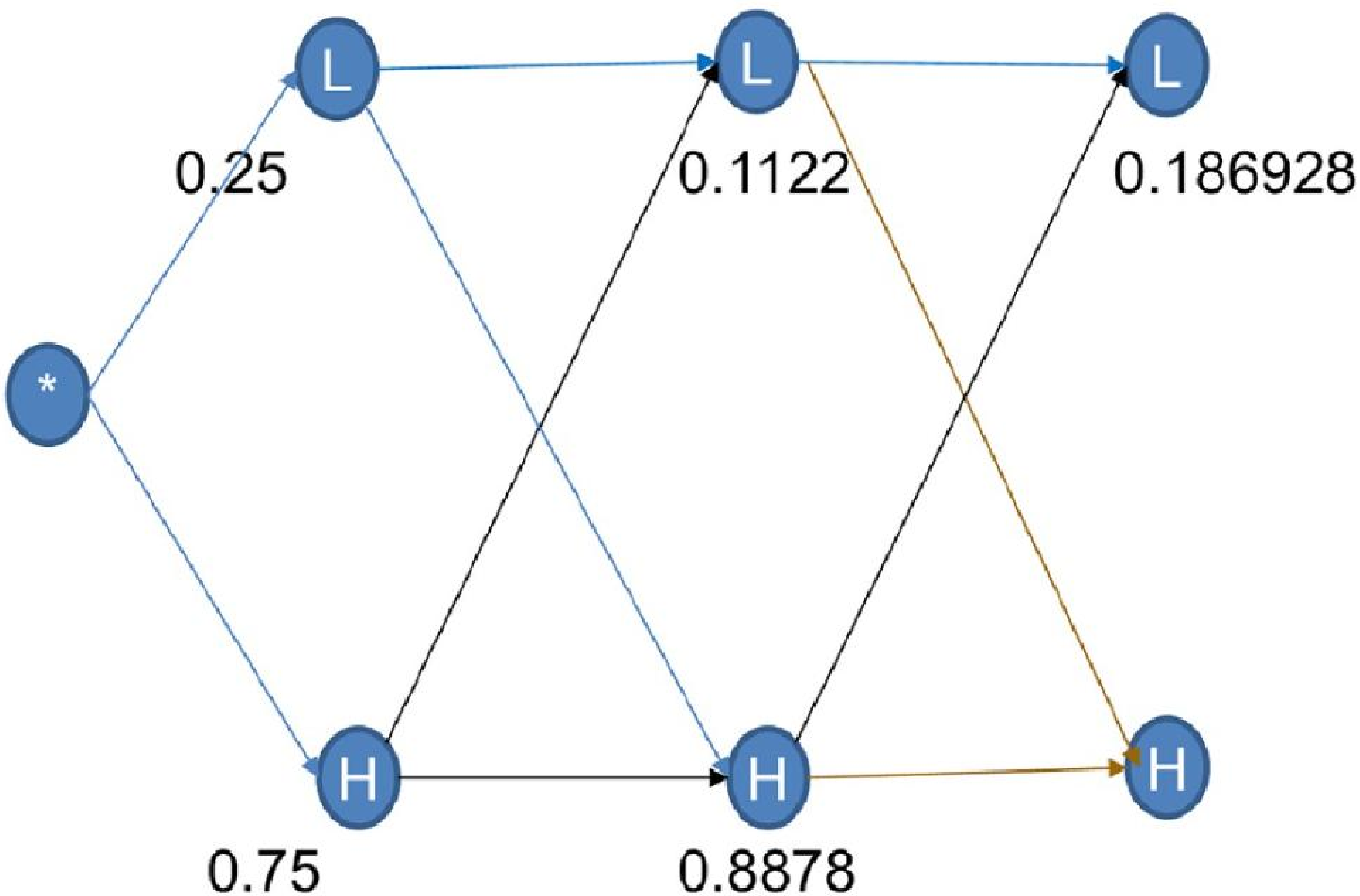
$P(U_{t-1} = HP)$	$P(U_{t-1} = LP)$	←Previous
0.2	0.5	$P(U_t = LP)$
0.8	0.5	$P(U_t = HP)$

Evidence / Sensor Model/ Emission Probability Matrix

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

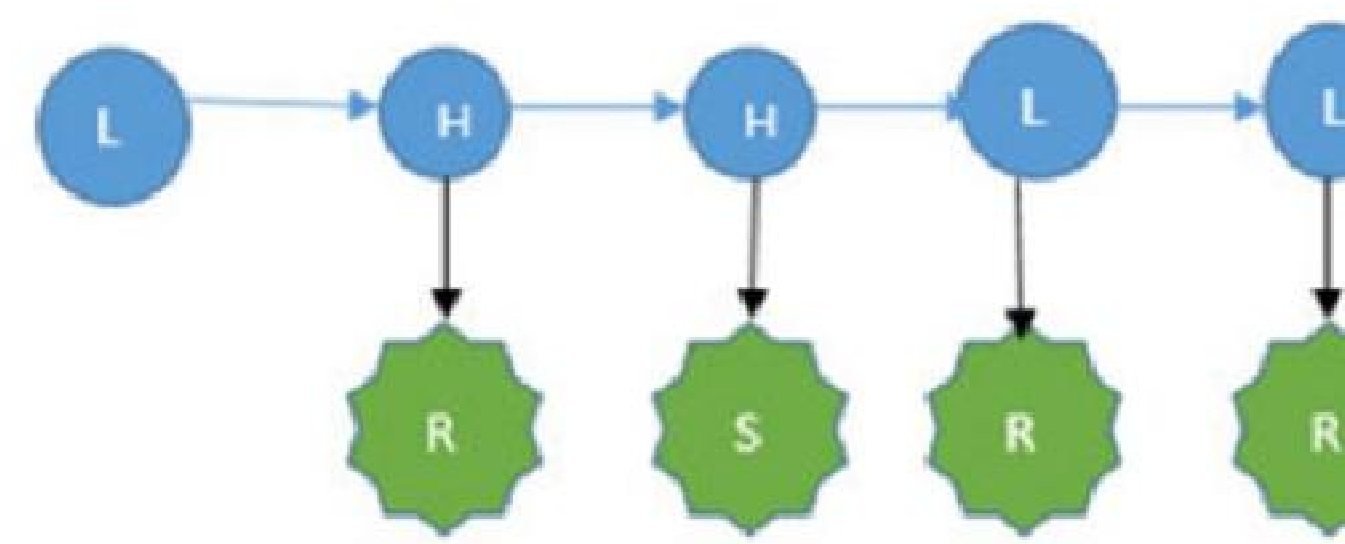
Hidden Morkov Model

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	$P(U_t = HP)$

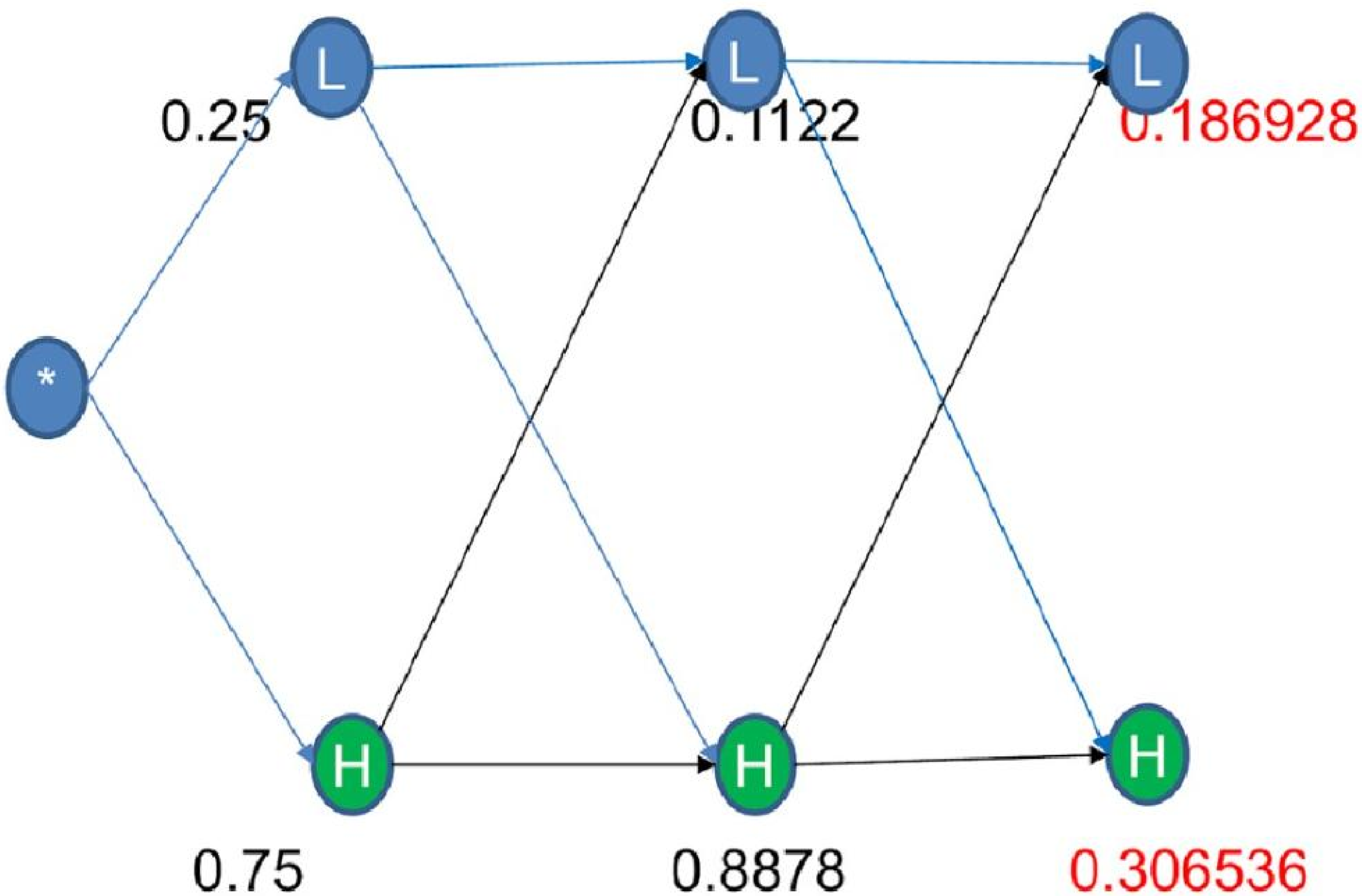
Evidence / Sensor Model/ Emission Probability Matrix

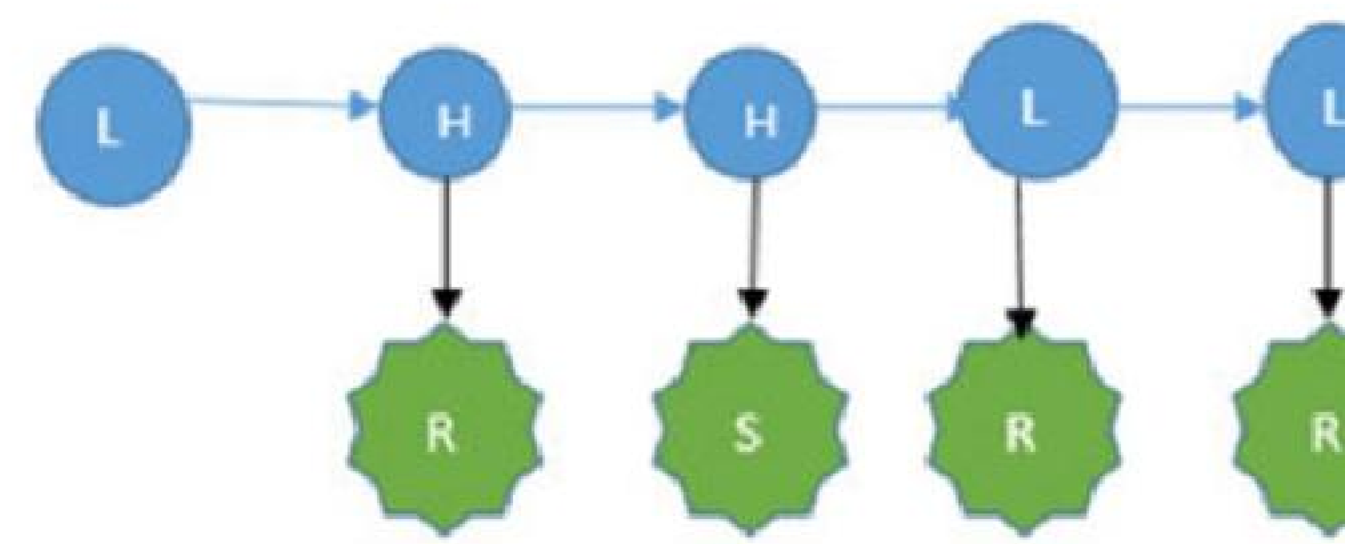
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

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

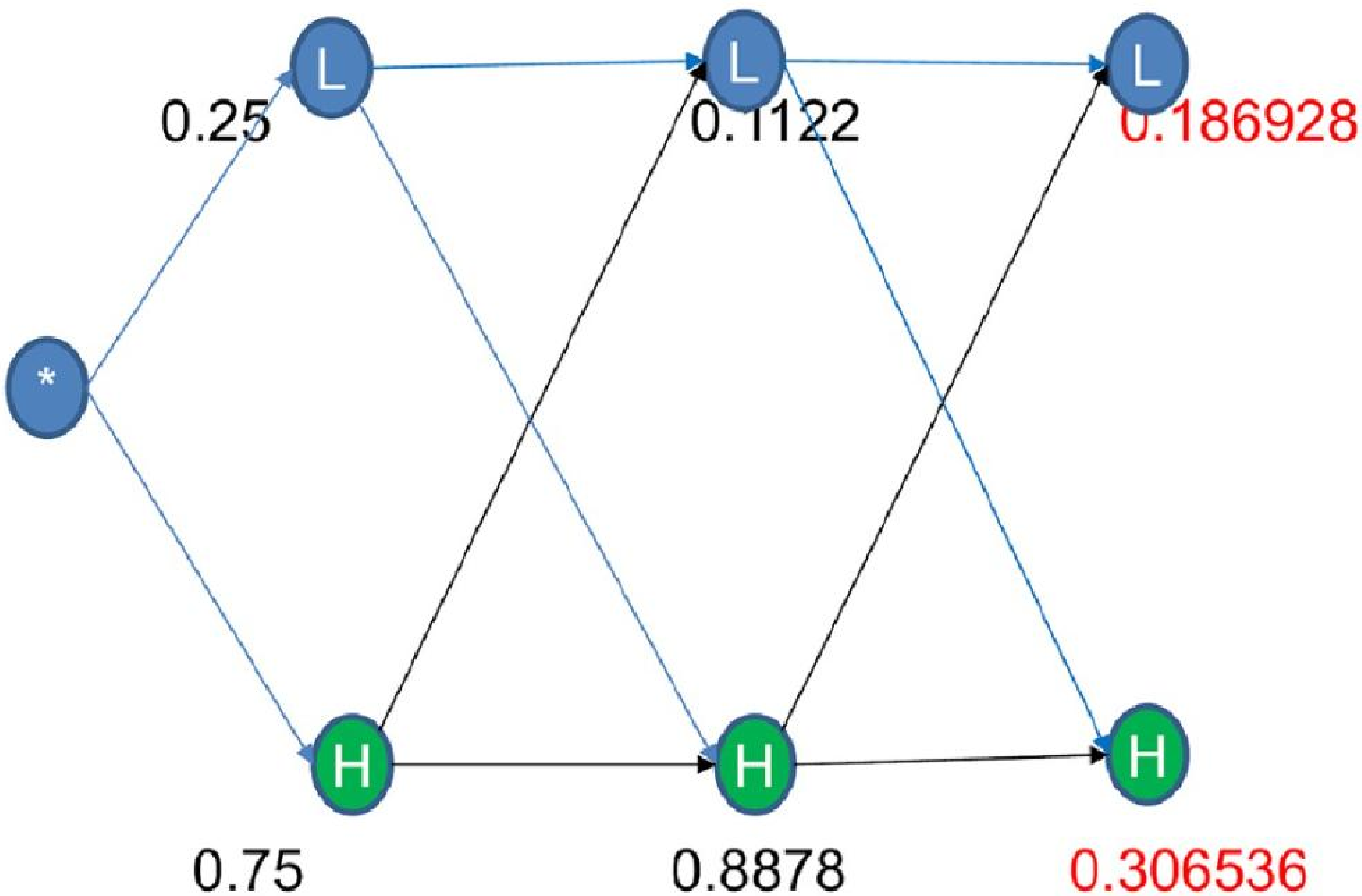
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

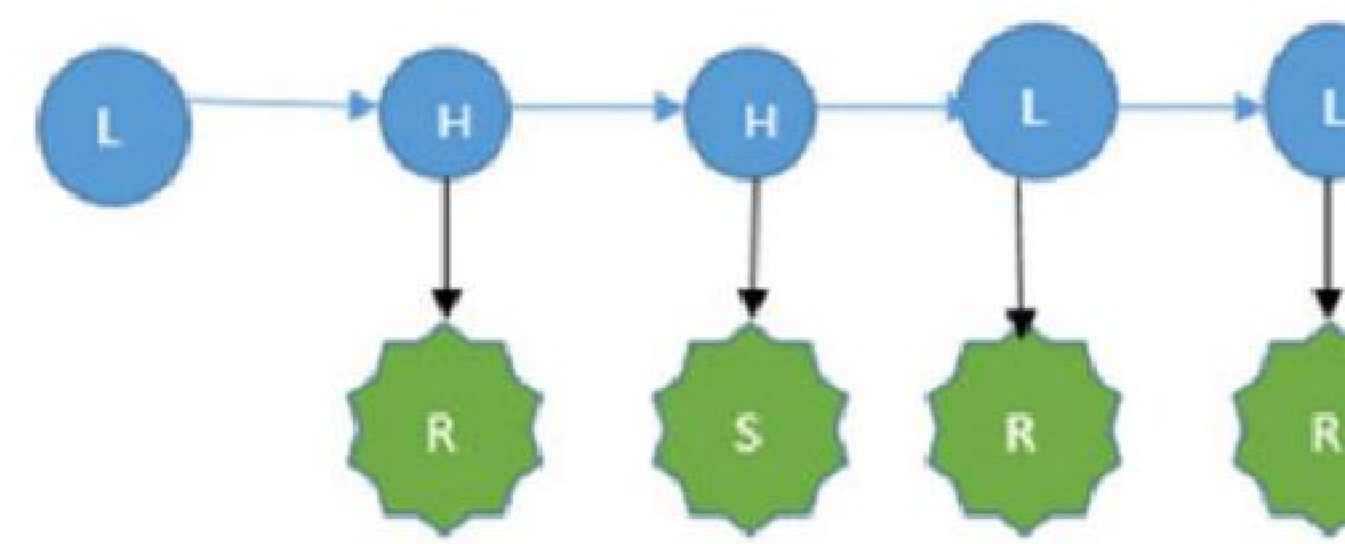
Hidden Morkov Model

Forward Propagation Algorithm : S-S-R

Termination Phase:

(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	$P(E_t = \text{Rainy})$
0.2	0.6	$P(E_t = \text{Sunny})$

Hidden Morkov Model

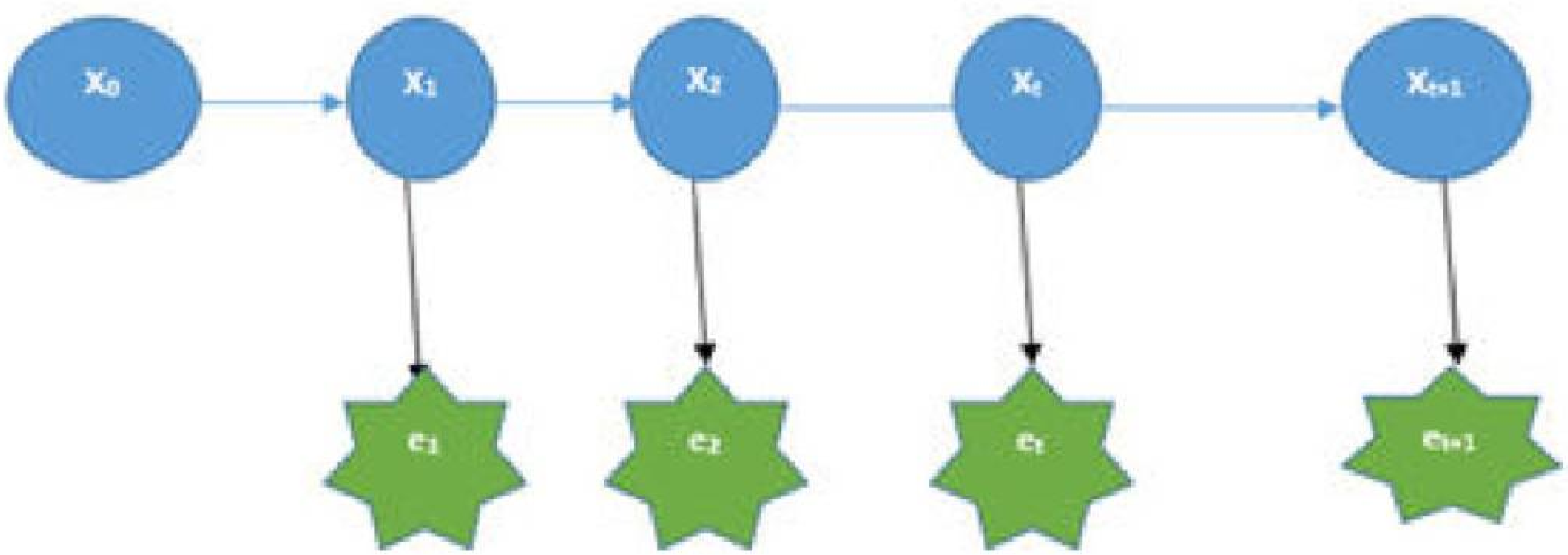
Inference: Type -3

Filtering : Forward Propagation Algorithm

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

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

$$P(Xt_{+1} | E_{1\ldots t+1}) = \boldsymbol{\alpha}\, P(e_{t+1} | Xt_{+1}) * \sum_{X_t} P(X_{t+1}|X_t) * \boldsymbol{P}(\mathbf{X_t}| \mathbf{E_{1..t}})$$



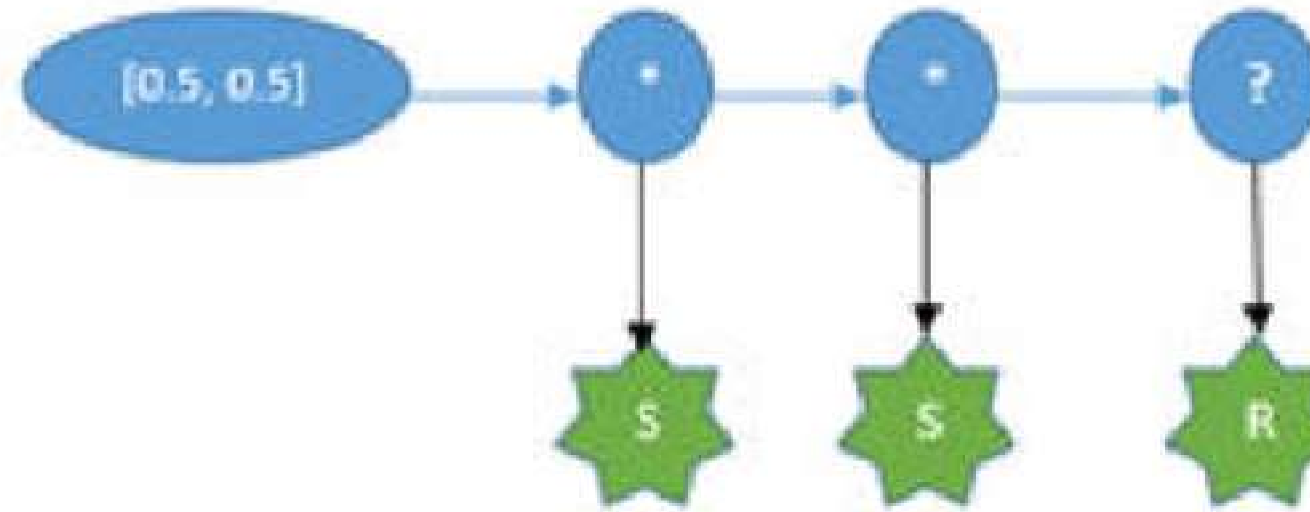
Transition Model / Probability Matrix

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

Evidence / Sensor Model/ Emission Probability Ma

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

Hidden Morkov Model



-3

Filtering : Forward Propagation Algorithm

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

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

$P(X_3 \mid SSR) = P(X_3 \mid S,S,R)$

$$= \frac{\boldsymbol{P(R\,|X_3\, S,S)} * P(X_3|\, S,S)}{P(R)}$$

$$= \frac{P(R\,|X_3) * \boldsymbol{P(X_3|\, S,S)}}{P(R)}$$

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

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

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	$P(U_t = HP)$

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

Evidence / Sensor Model/ Emission Probability Ma

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

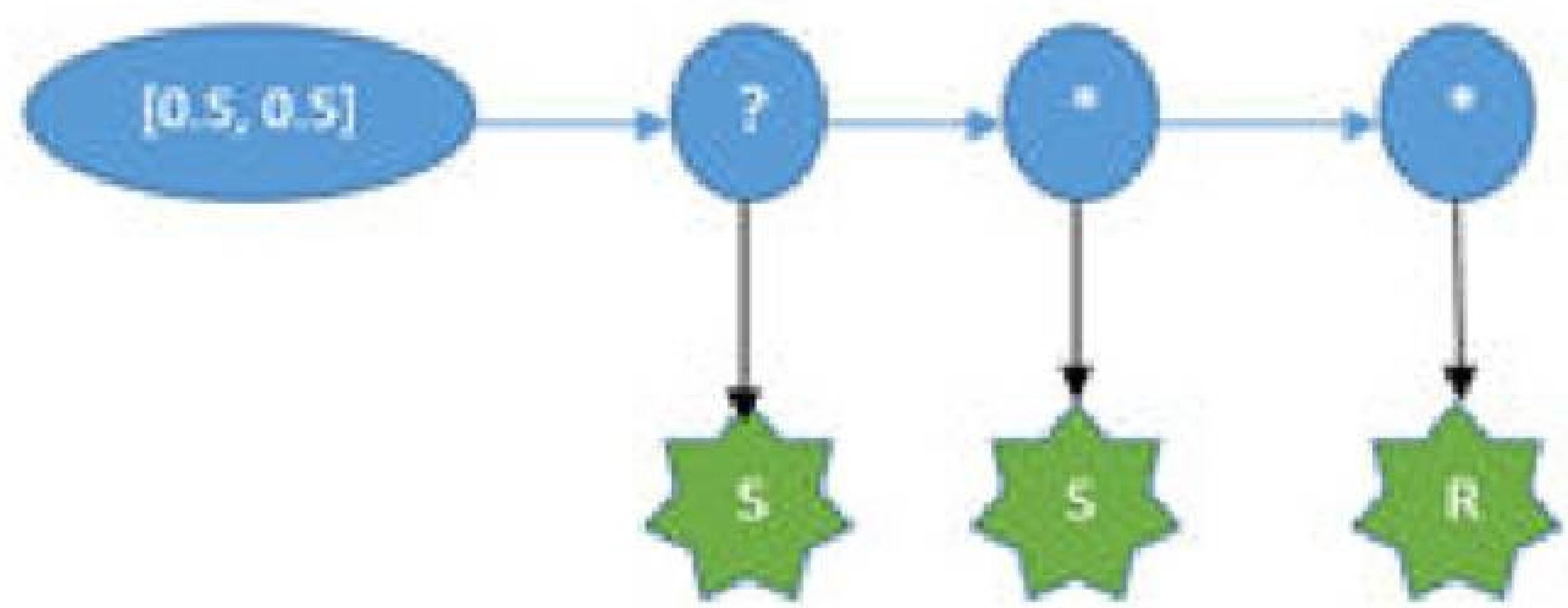
Hidden Morkov Model

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



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

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

$P(X_1 \mid SSR) = P(X_1 \mid S,S,R)$

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

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

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

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

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

Transition Model / Probability Matrix

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

Evidence / Sensor Model/ Emission Pr

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

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

Hidden Morkov Model

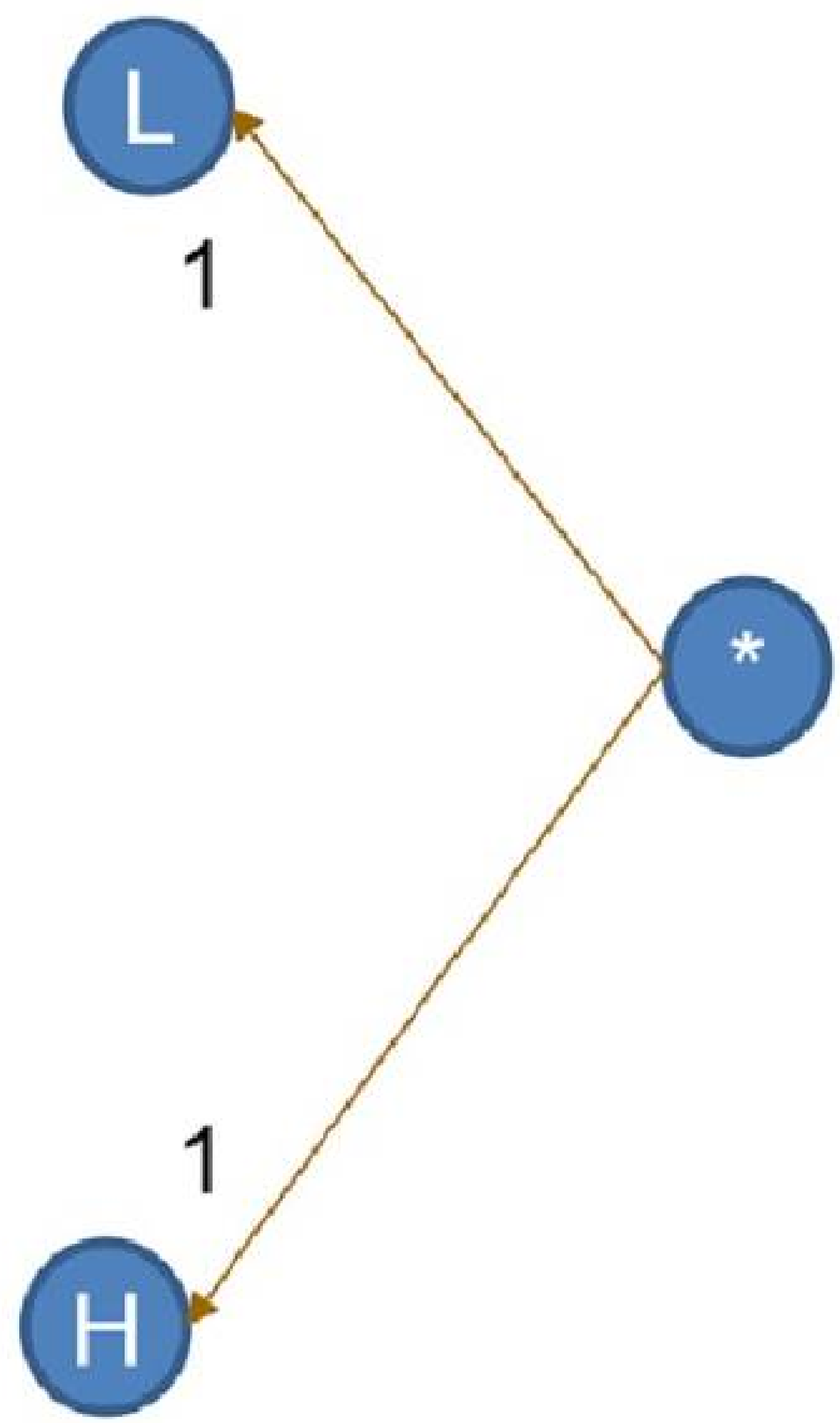
Backward Propagation Algorithm

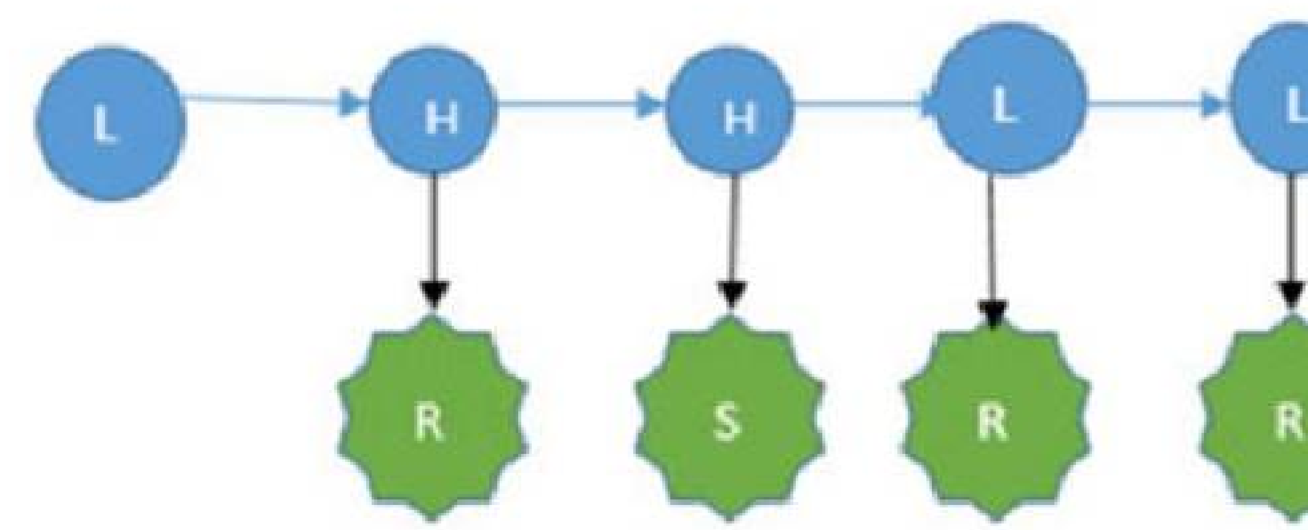
Pressure sequence observation: **S-S-R**

<u>Initialization Phase: </u> 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





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)

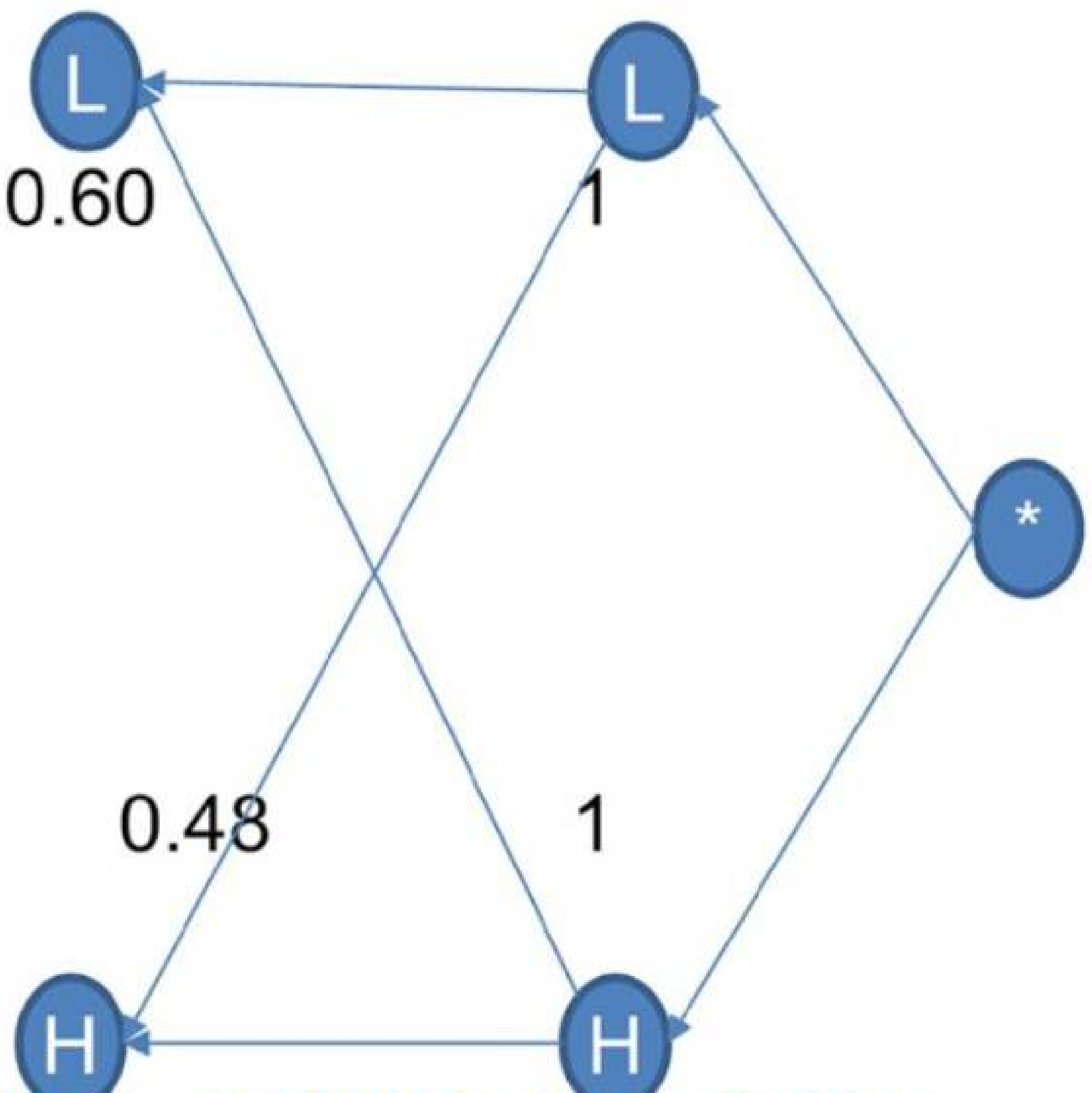
P(L|H)*P(R|L)*P(.|L) = 0.2*0.8 * 1= 0.16

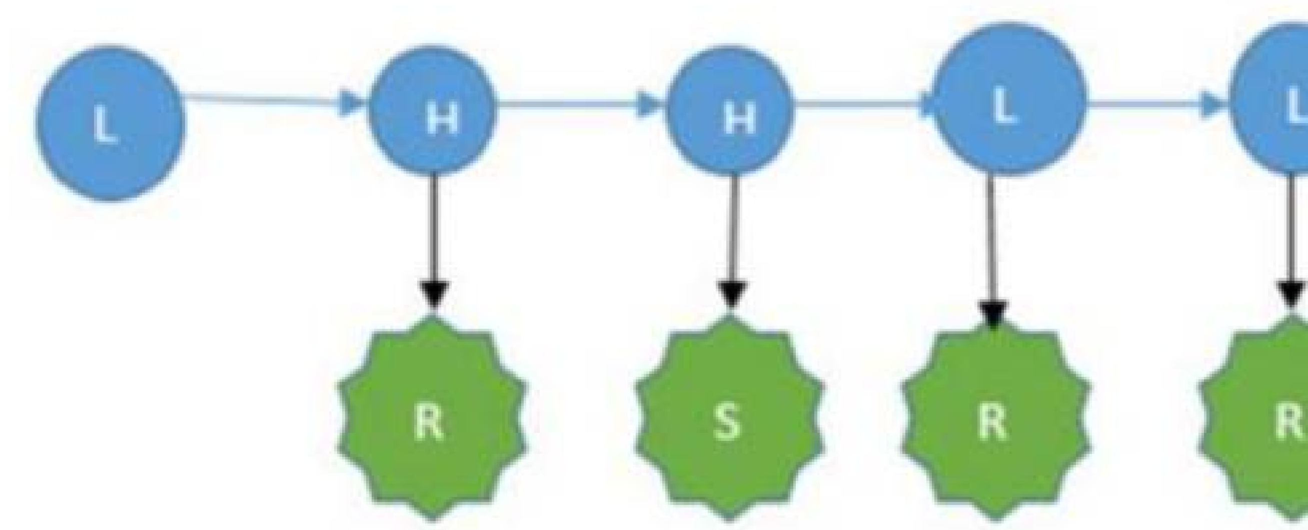
P(H|H)*P(R|H)*P(.|H) = 0.8*0.4 *1= 0.32

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





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

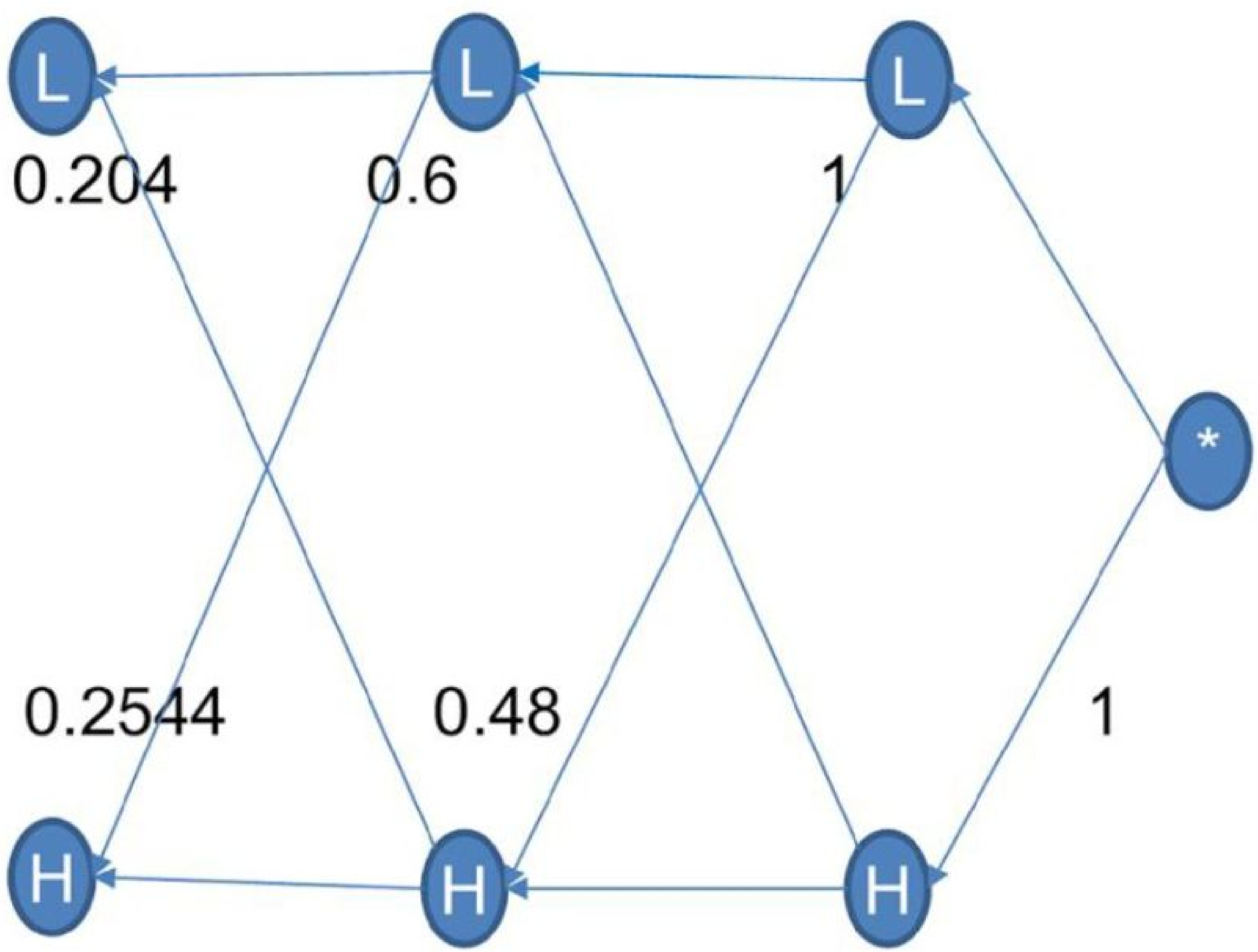
P(X_t = LP)	P(X_t = HP)	←Unobserved Evidence v
0.8	0.4	P(E_t = Rainy)
0.2	0.6	P(E_t = Sunny)

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

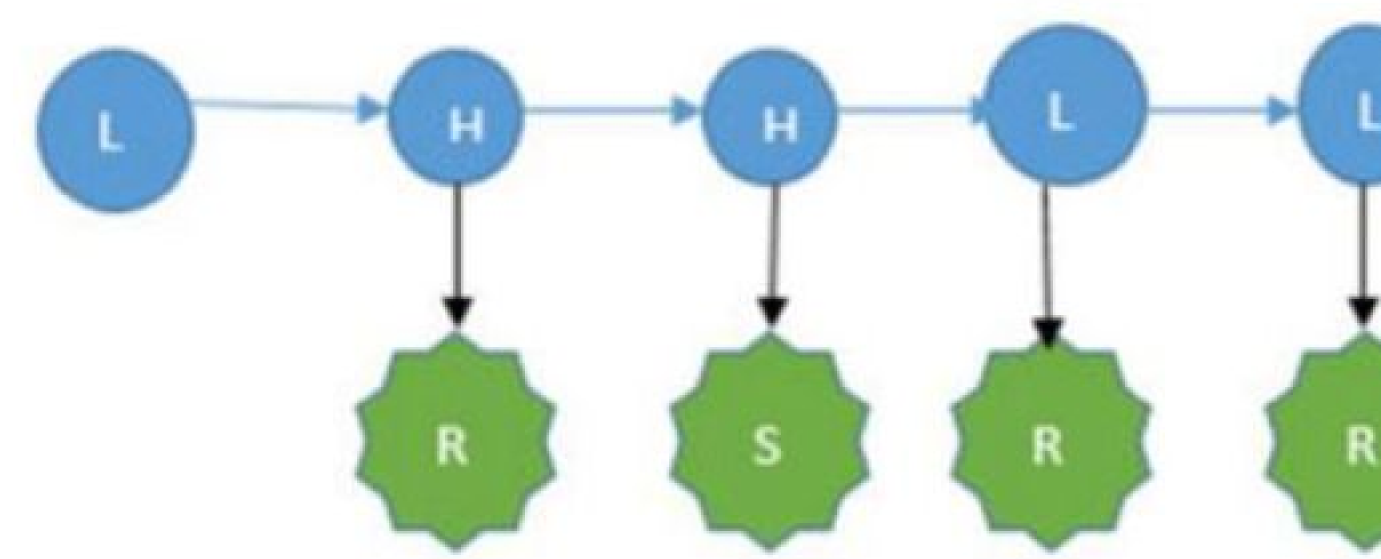
Recursion Phase:

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	$P(U_t = HP)$

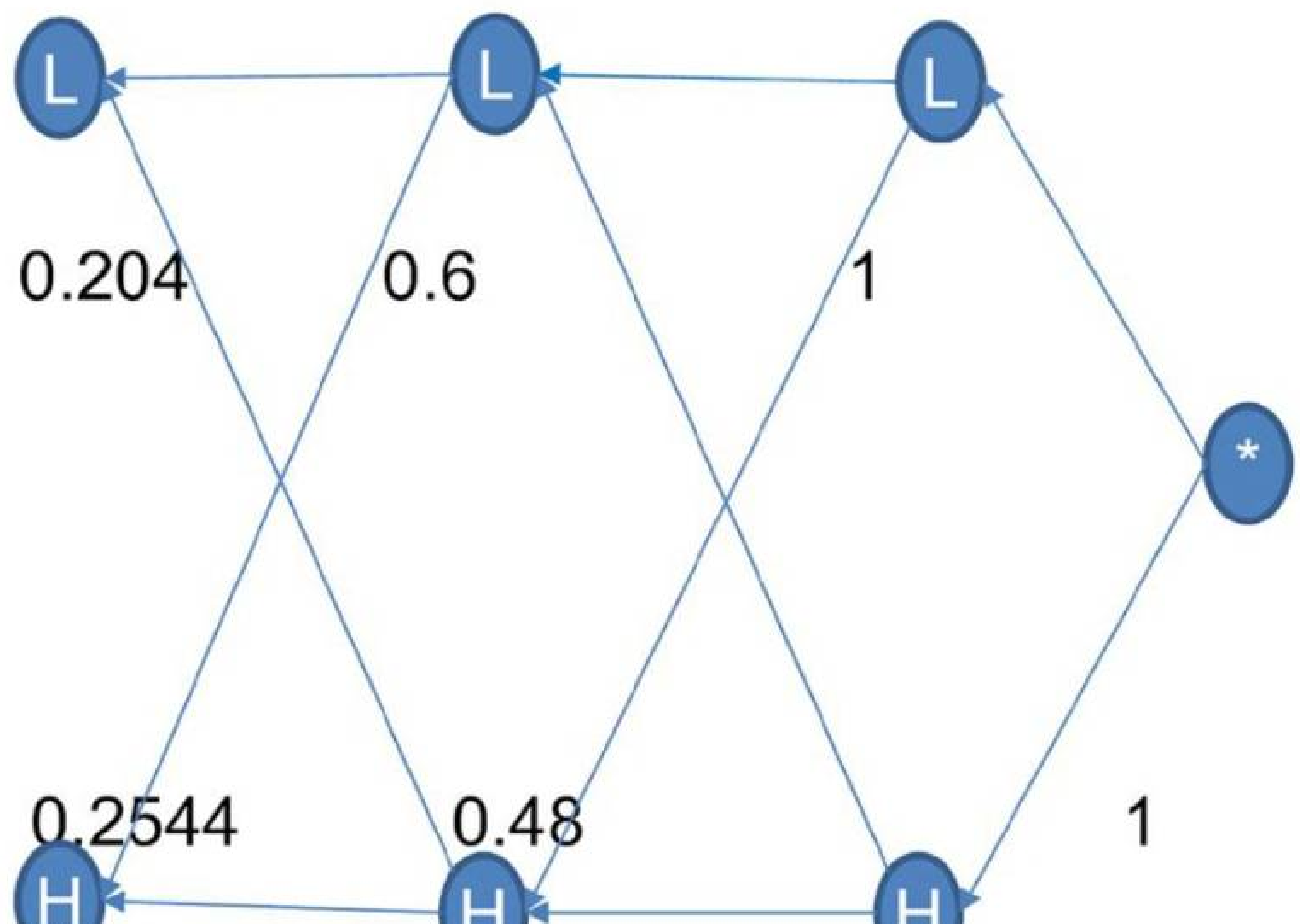
Evidence / Sensor Model/ Emission Probability Matrix

$P(X_t = LP)$	$P(X_t = HP)$	←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

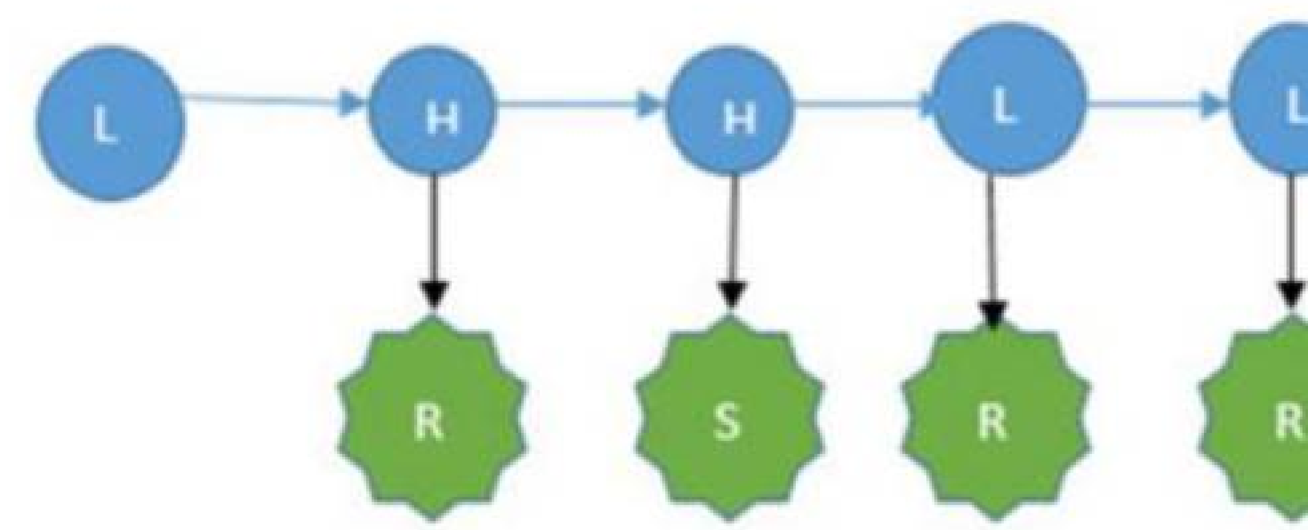
P(L)*P(S|L)*MSG(L`) = 0.5*0.2 * 0.204= 0.0204



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

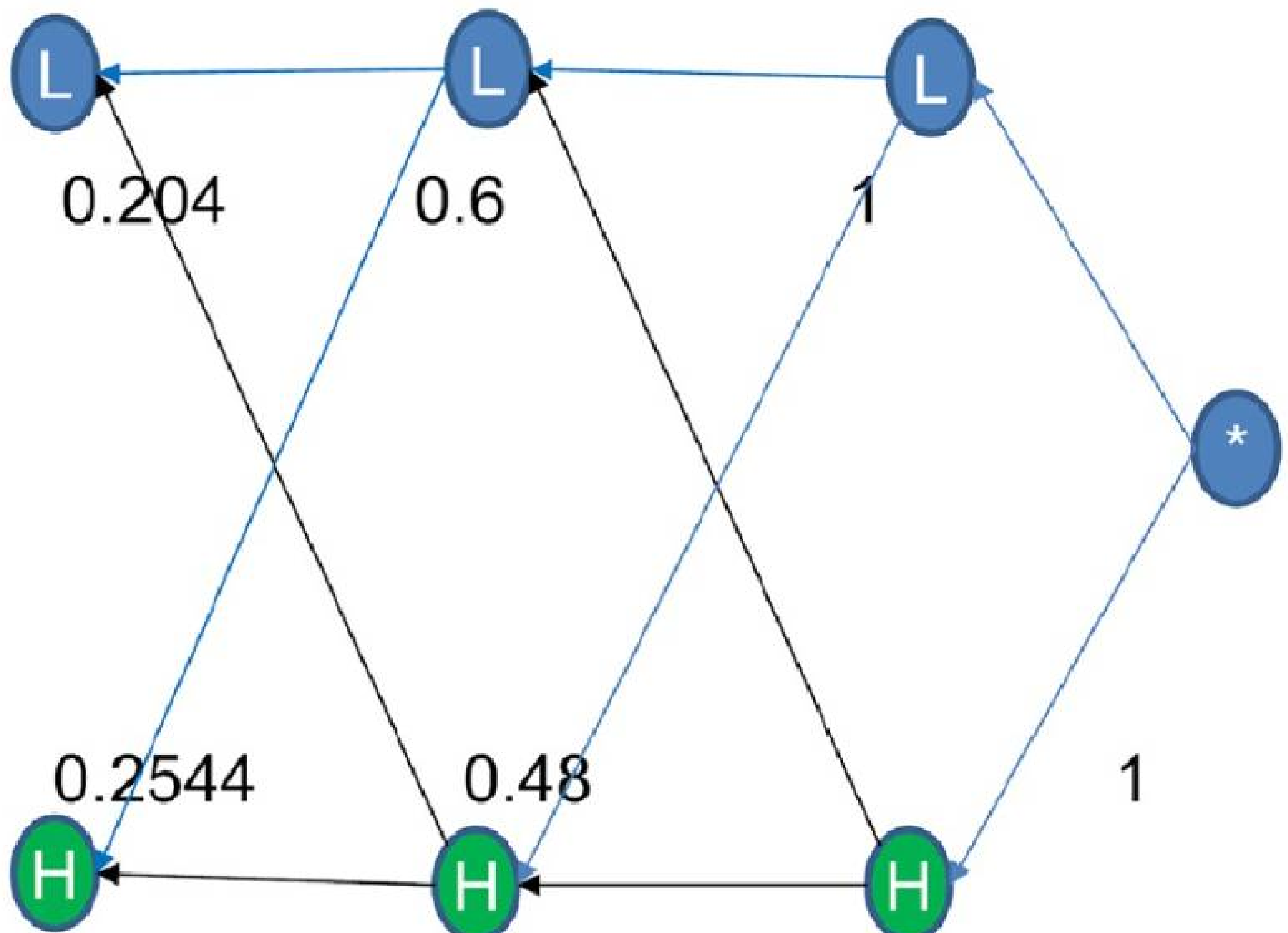
$P(X_t = LP)$	$P(X_t = HP)$	←Unobserved Evidence v
0.8	0.4	$P(E_t = Rainy)$
0.2	0.6	$P(E_t = Sunny)$

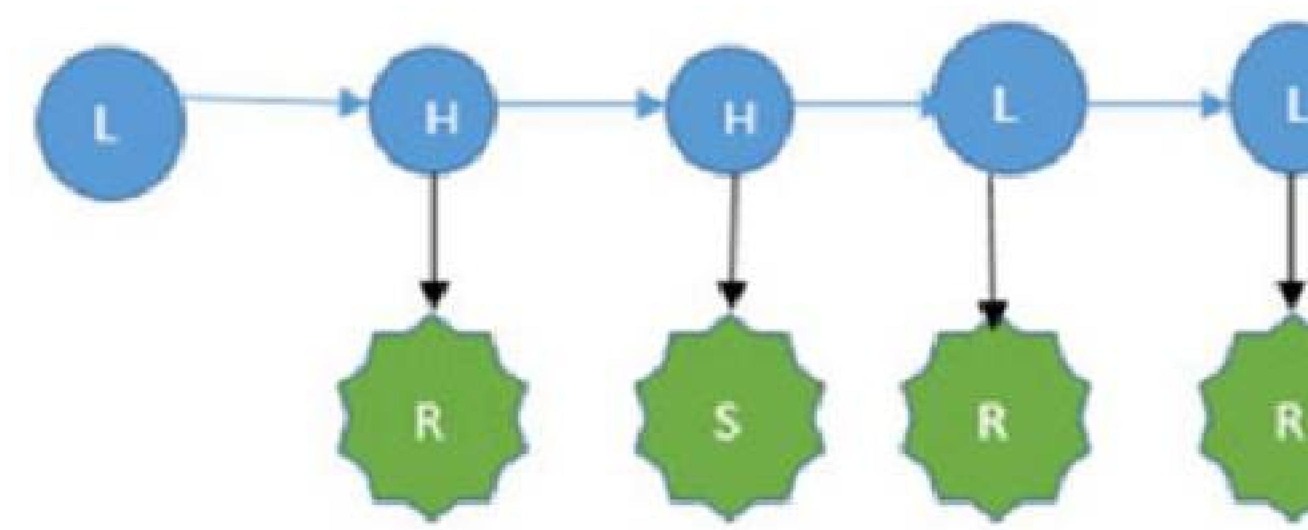
Hidden Morkov Model

Forward Backward Propagation Algorithm : S-S-R

P(X2 | SSR) = α * P(X2|SS) *P(R|X2)

P(X2 | SSR) = α * (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

$\alpha_t(j) \quad = \alpha_{t-1}(i)\ a_{i,j}\ b_j(o_t)$

i

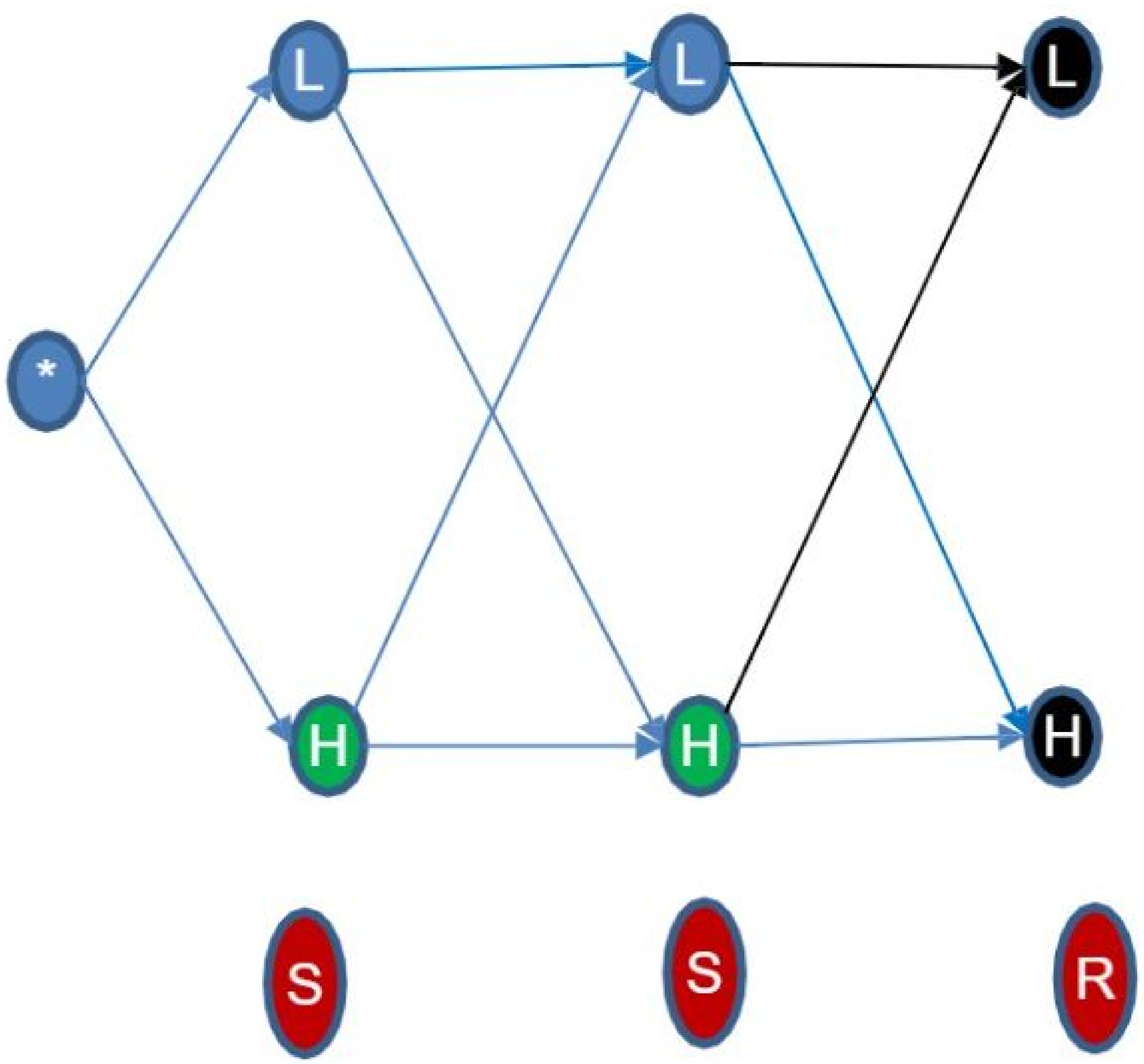
$P(O_{1..t} \mid \boldsymbol{\lambda})$

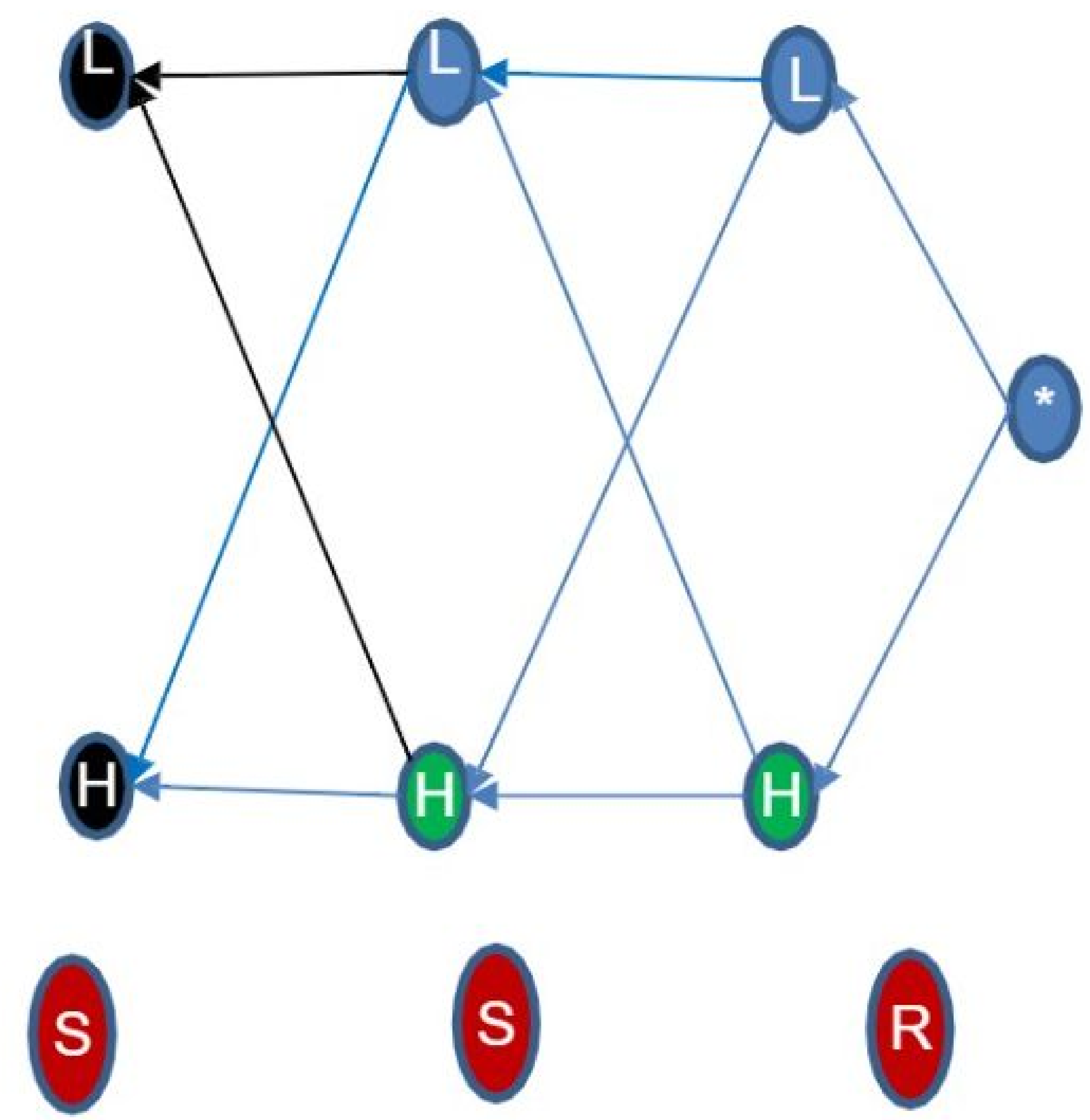
Backward Path Probability

$\beta_t(i) = {}_{j} \quad \beta_{t+1}(j)\ a_{i,j}\ b_j(o_{t+1})$

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

$\boldsymbol{\gamma}_t(i) = P(X_t \mid O_{1\ldots t\ \ldots t+1,\, t+2..t+k} \mid \boldsymbol{\lambda})$: Forward – Backward Algorithm





function FORWARD-BACKWARD(**ev**, *prior*) **returns** a vector of probability distributions
  inputs: **ev**, a vector of evidence values for steps $1, \ldots, t$
    prior, the prior distribution on the initial state, $\mathbf{P}(\mathbf{X}_0)$
  local variables: **fv**, a vector of forward messages for steps $0, \ldots, t$
    b, a representation of the backward message, initially all 1s
    sv, a vector of smoothed estimates for steps $1, \ldots, t$

  $\mathbf{fv}[0] \leftarrow prior$
  for $i = 1$ **to** t **do**
    $\mathbf{fv}[i] \leftarrow$ FORWARD($\mathbf{fv}[i-1], \mathbf{ev}[i]$)
  for $i = t$ **downto** 1 **do**
    $\mathbf{sv}[i] \leftarrow$ NORMALIZE($\mathbf{fv}[i] \times \mathbf{b}$)
    $\mathbf{b} \leftarrow$ BACKWARD($\mathbf{b}, \mathbf{ev}[i]$)
  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

Text & Natural Language Processing

HMM Application

Initial	Prob
Noun	
aDverb	
Verb	
adJective	
Preposition	
Determiner	

N	D	V	J	P	
					N
					D
					V
					J
					A
					P
					E

- 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 boy
  N V D P
- The tall tree is falling.
  D J N V V

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.

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

N	D	V	J	A	P	
0.25						Boy s
		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					Mor e
					1	Tha
		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

Init ial	Pro b
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

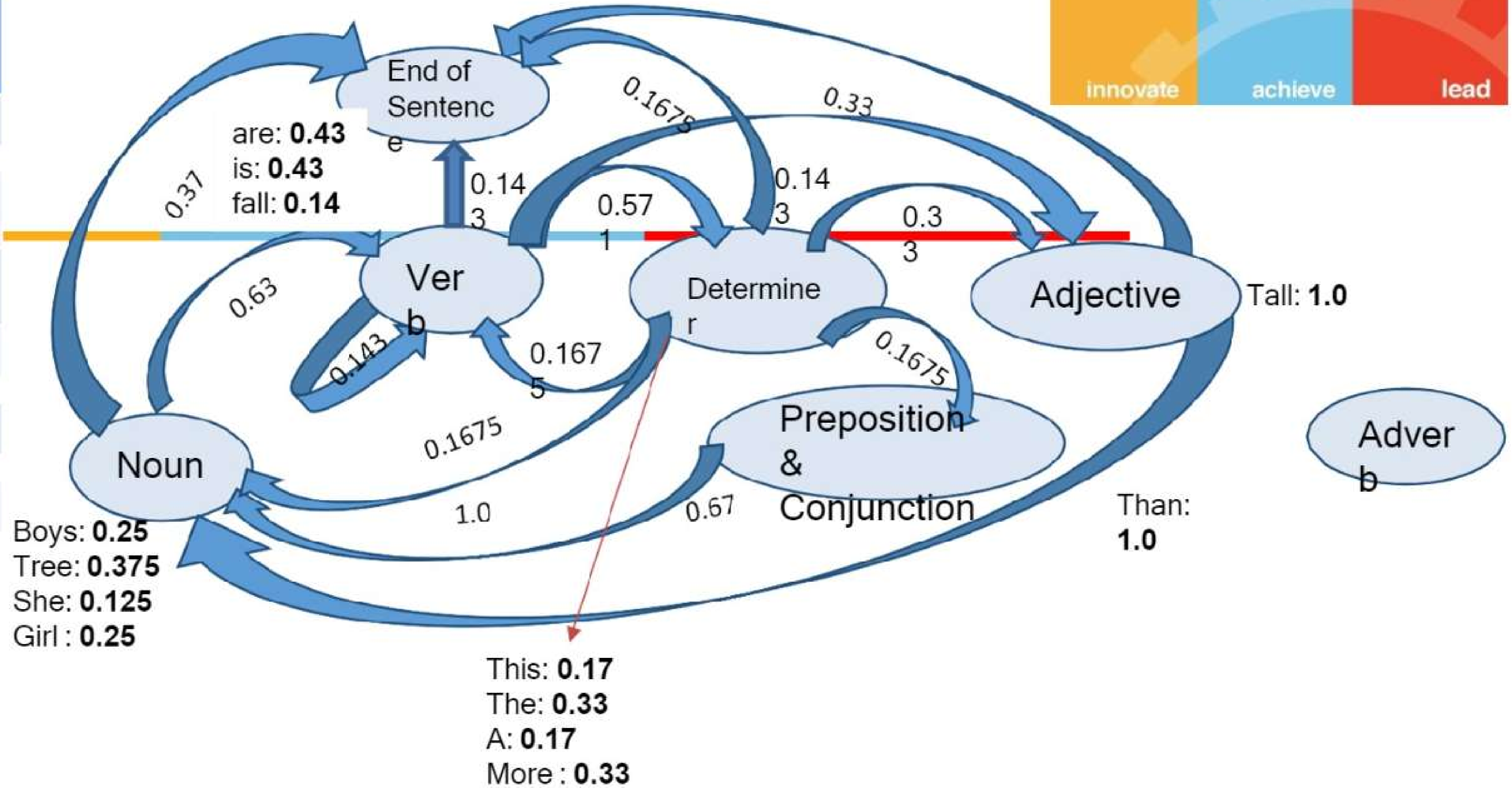
Exercise :

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

"Girls are falling"

N	D	V	J	A	P	
0.25						Boy s
		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					Mor e
					1	Tha n
		0.14				fall

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**

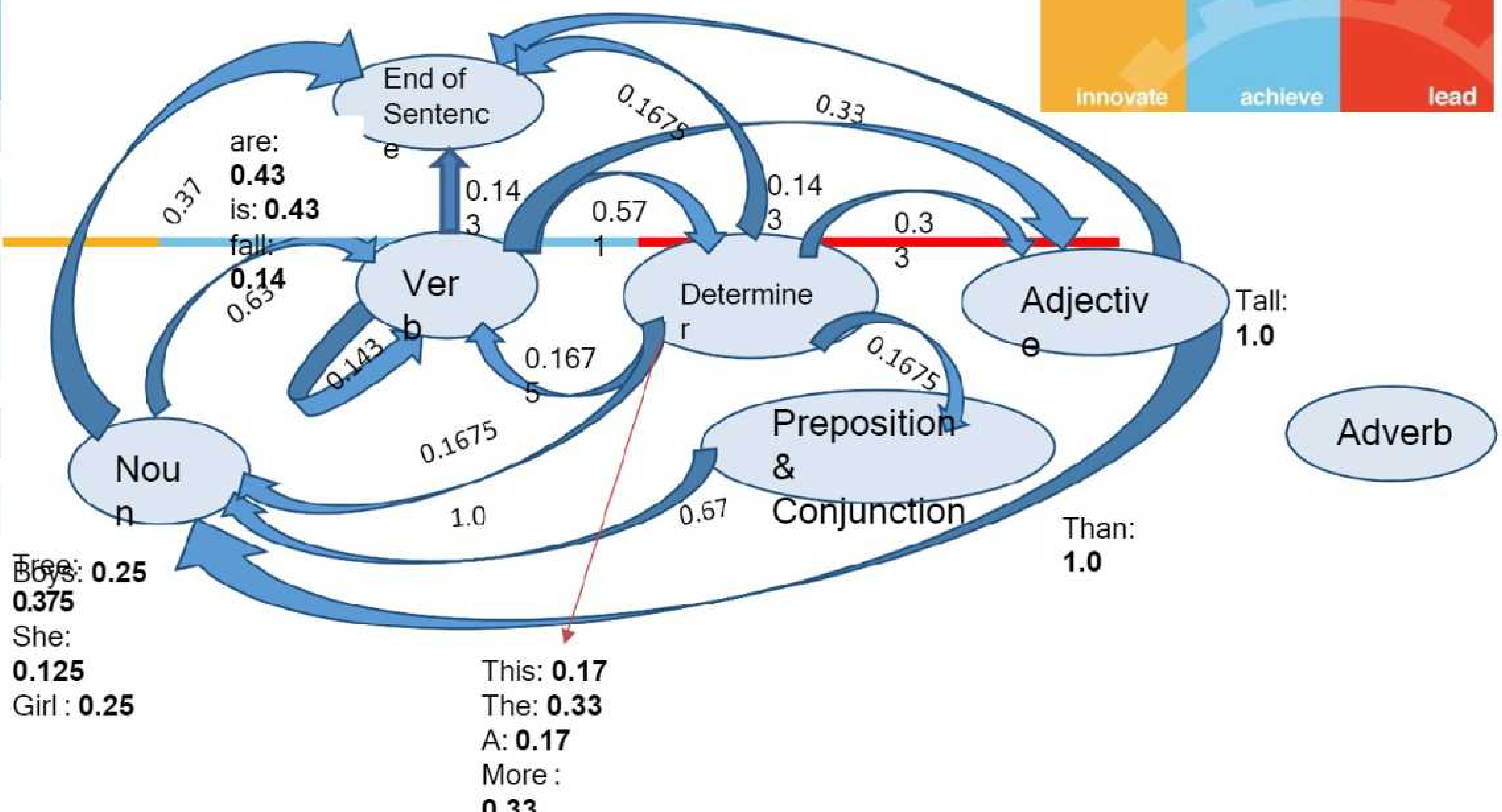
P(Girls , Verb) = P(Girl | Verb) * P(Verb | StartState) = 0 * 0 = 0 (Ideally with better corpus and the KB , for

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

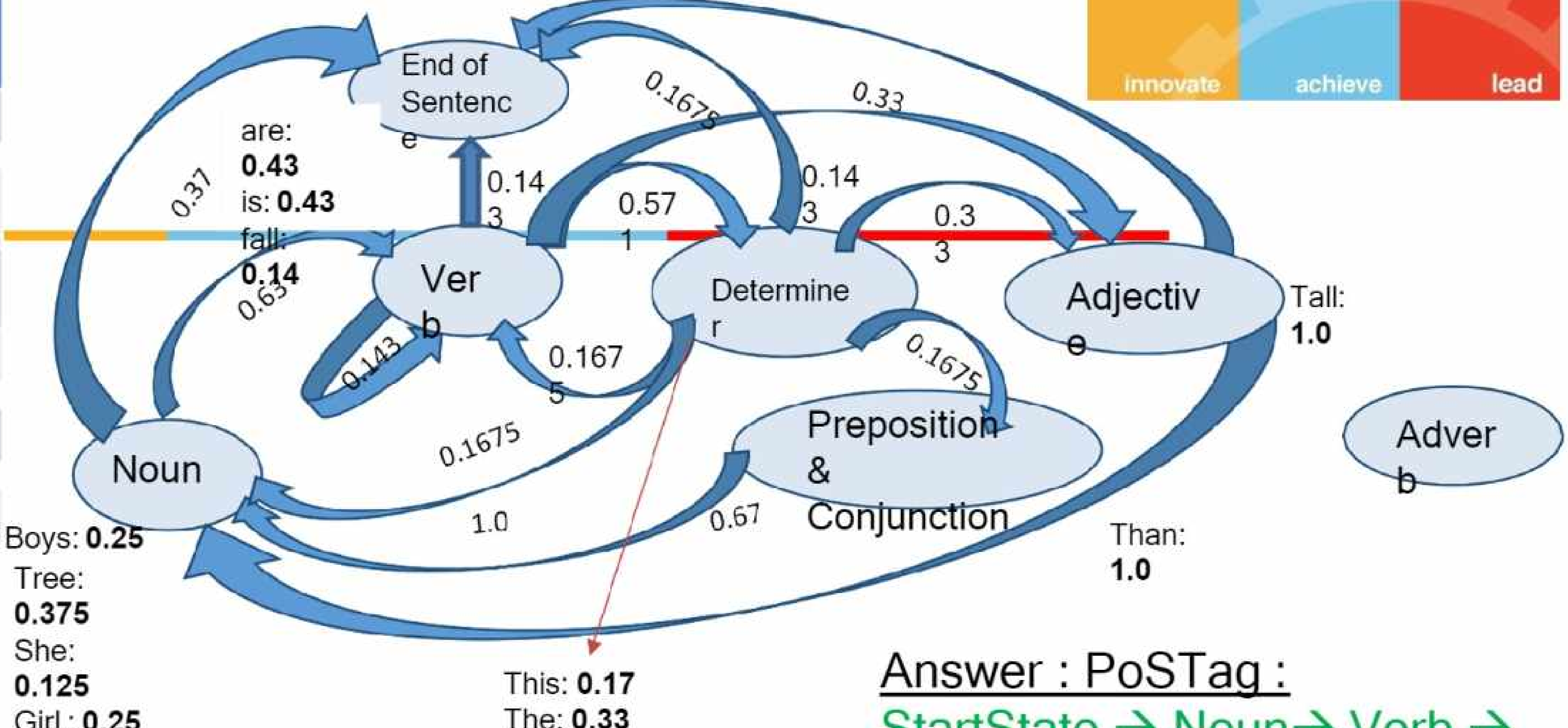
Initi	
N	0.67
D	0.33
V	0
J	0
A	0
P	0



"Girls are falling"

If Sequence is StartState → Noun	If Sequence is StartState → Verb
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	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"

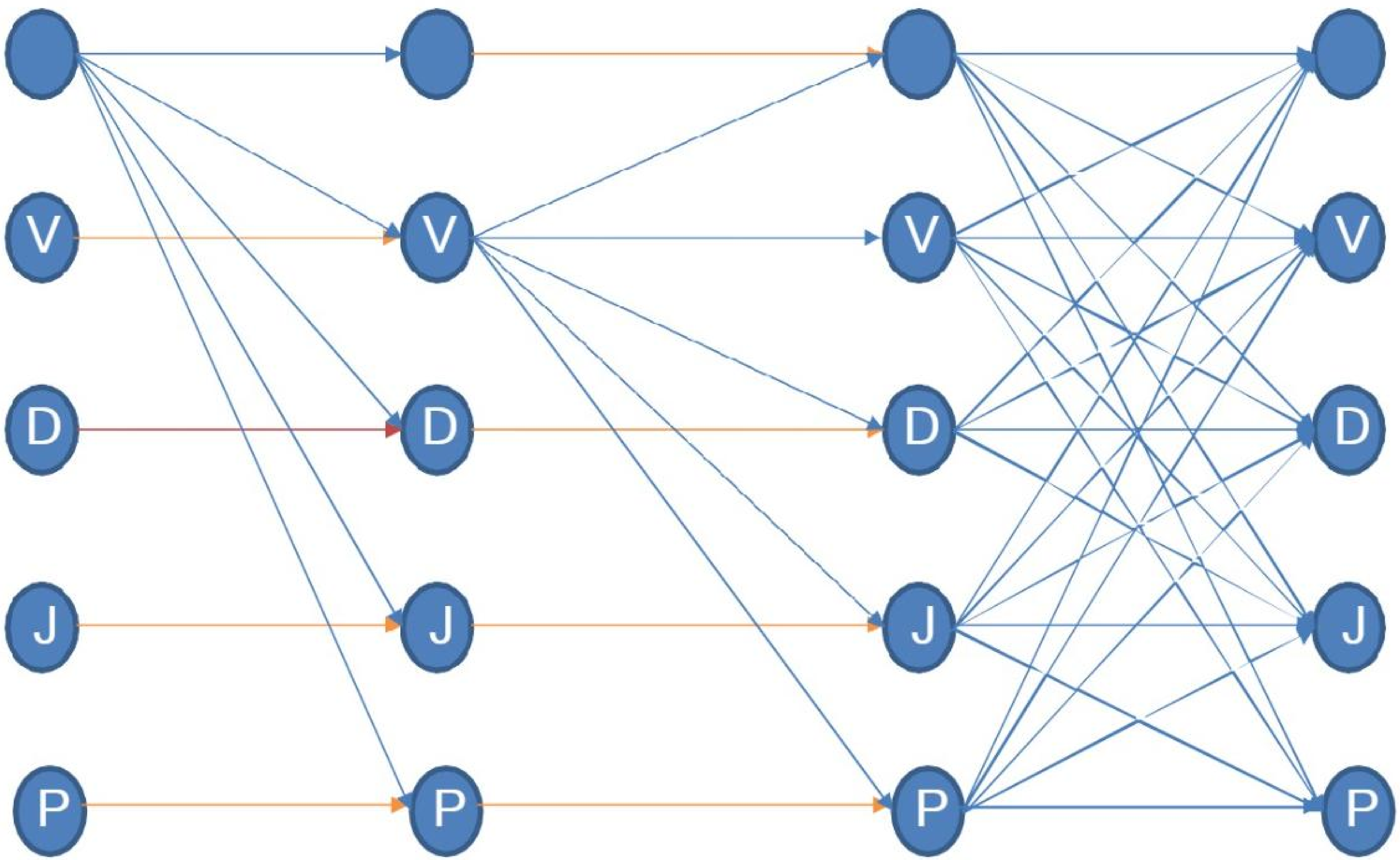
Answer : PoSTag :

StartState → Noun→ Verb → Verb

If Sequence is StartState→ Noun →Verb	If Sequence is StartState→ Verb ->Adjective
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	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 PARAMATERS:

{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

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

Evidence / Sensor Model/ Emission Pro

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserve 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 PARAMATERS:

{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

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

Evidence / Sensor Model/ Emission Pro

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserve 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 parAMATERS:

{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

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

Evidence / Sensor Model/ Emission Pro

$P(X_t = LP)$	$P(X_t = HP)$	←Unobserve 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 parAMATERS:

{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 Probab

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

HMM in Prevention of Network Security Threat

(Interesting Case Studies)

Hidden Morkov Model

Cyber Security

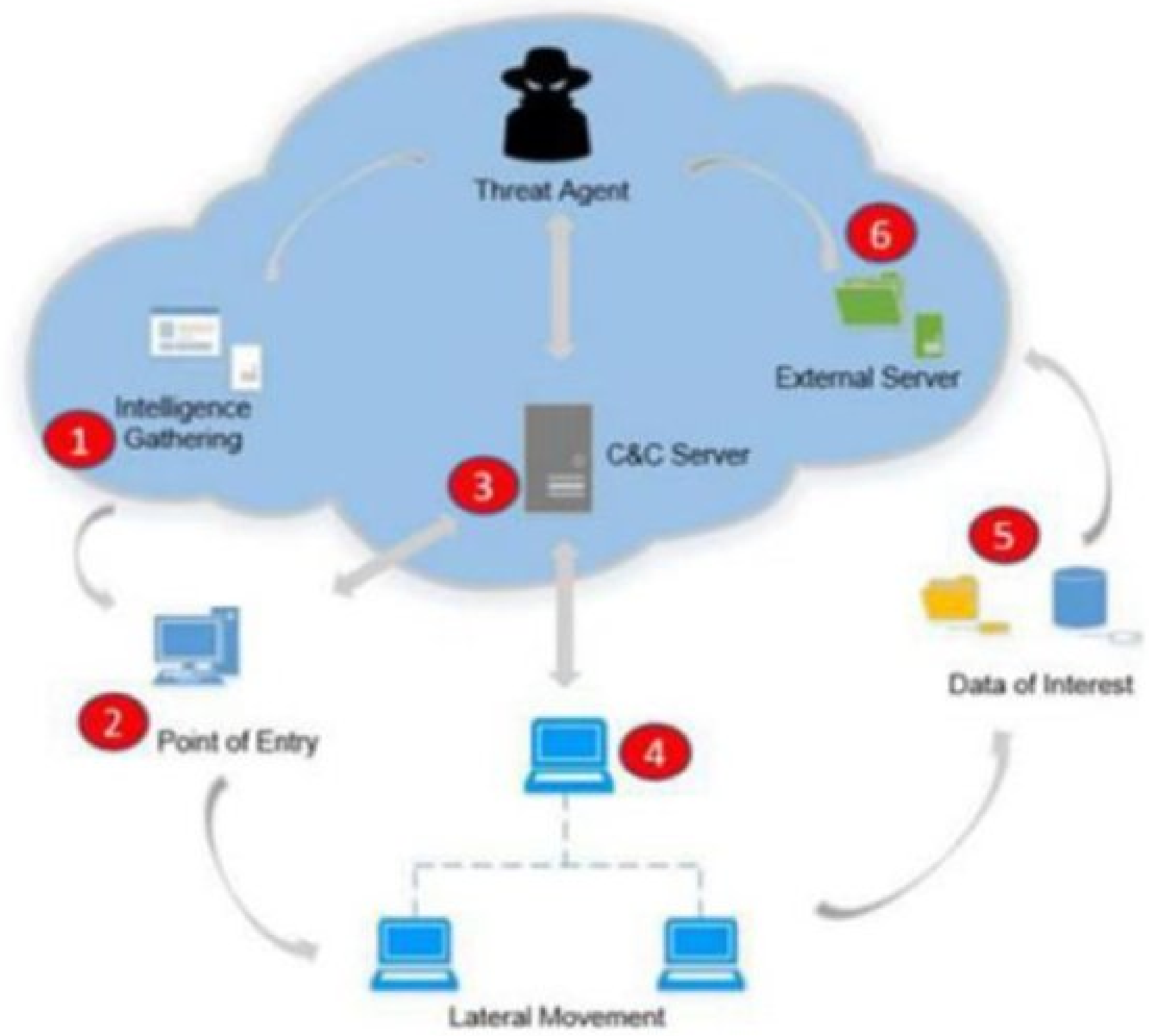
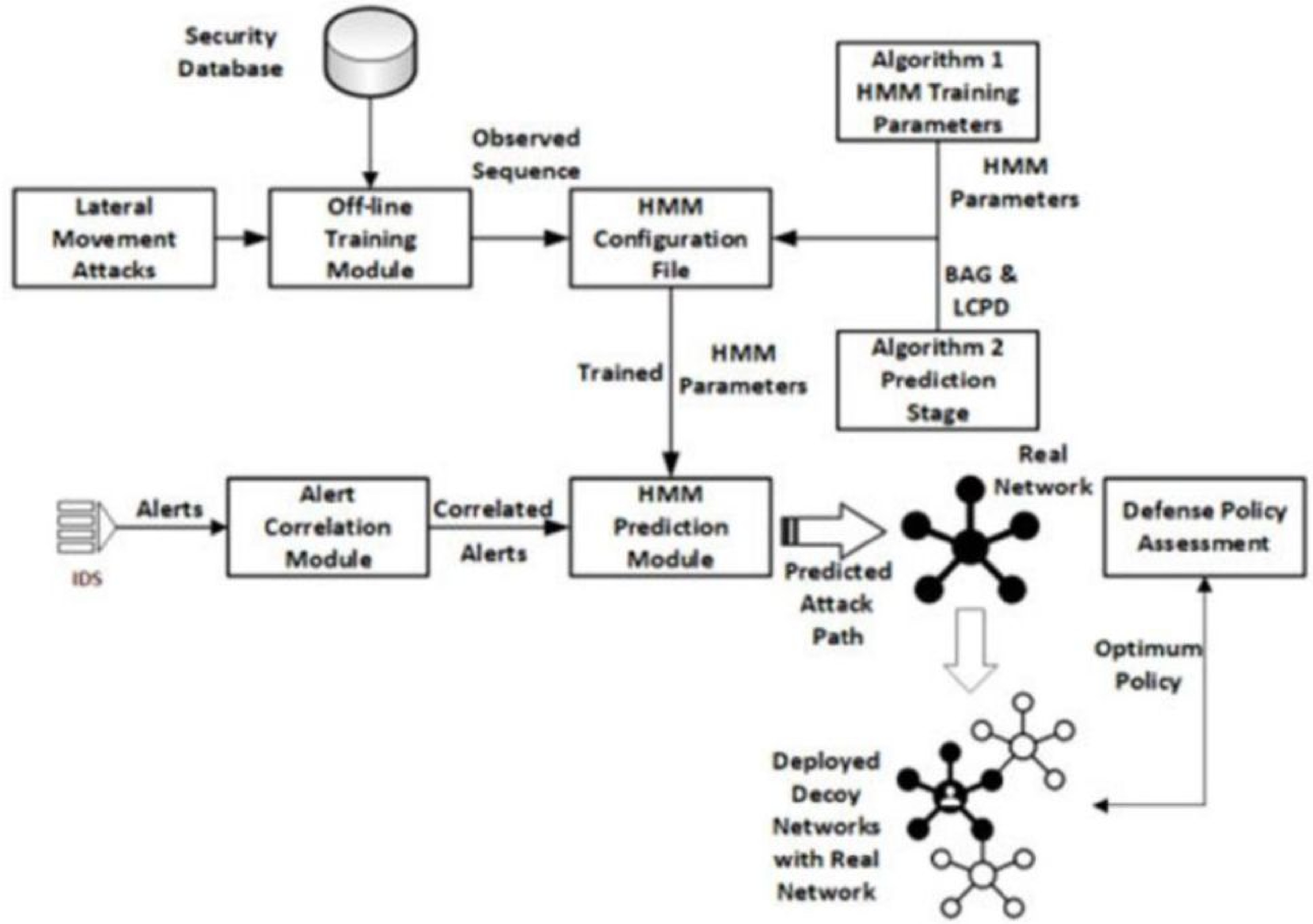


FIGURE 1. **Typical stages of APT attack.**

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

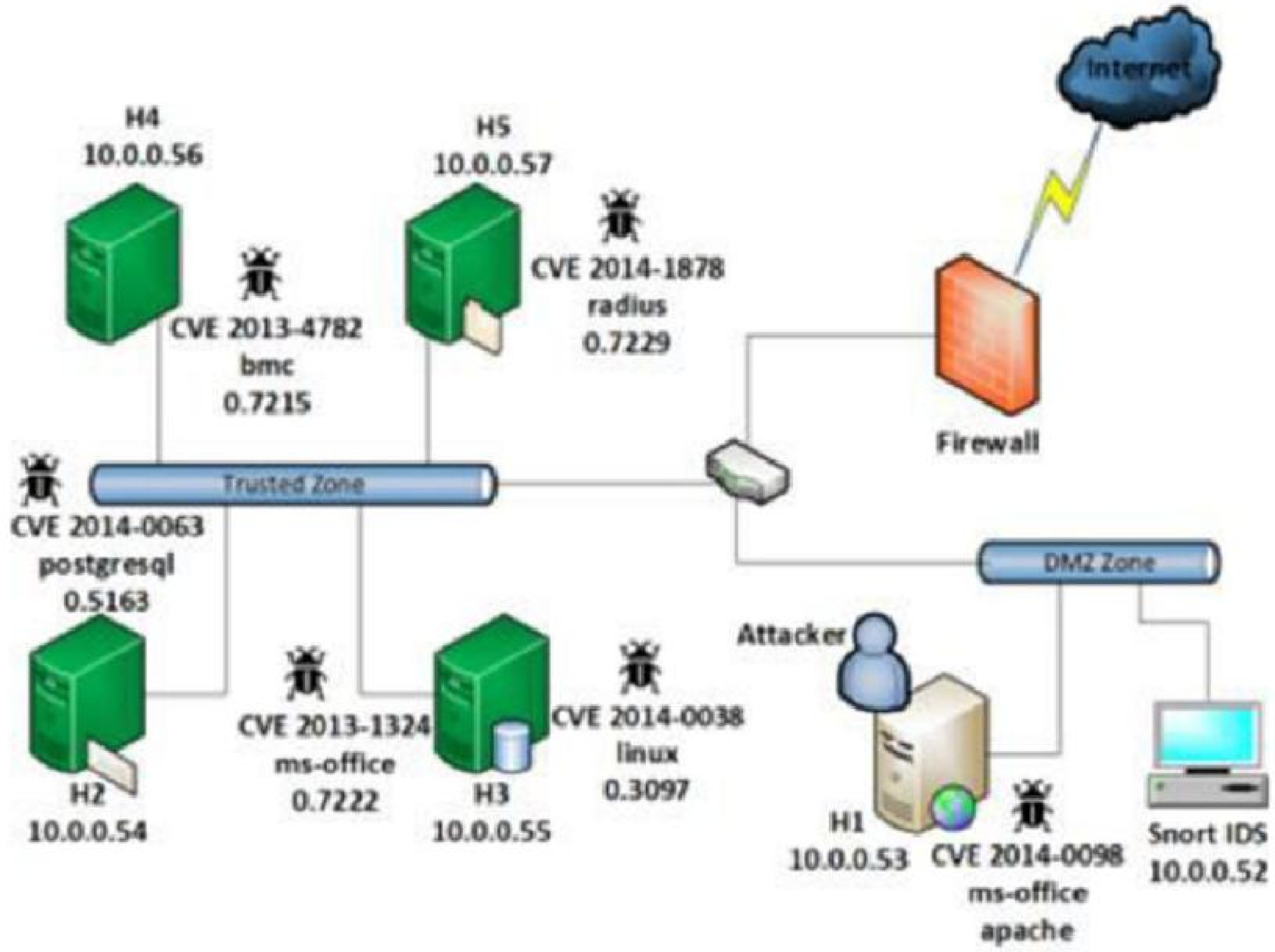


FIGURE 9. Experimental network topology.

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

Hidden Morkov Model

Cyber Security

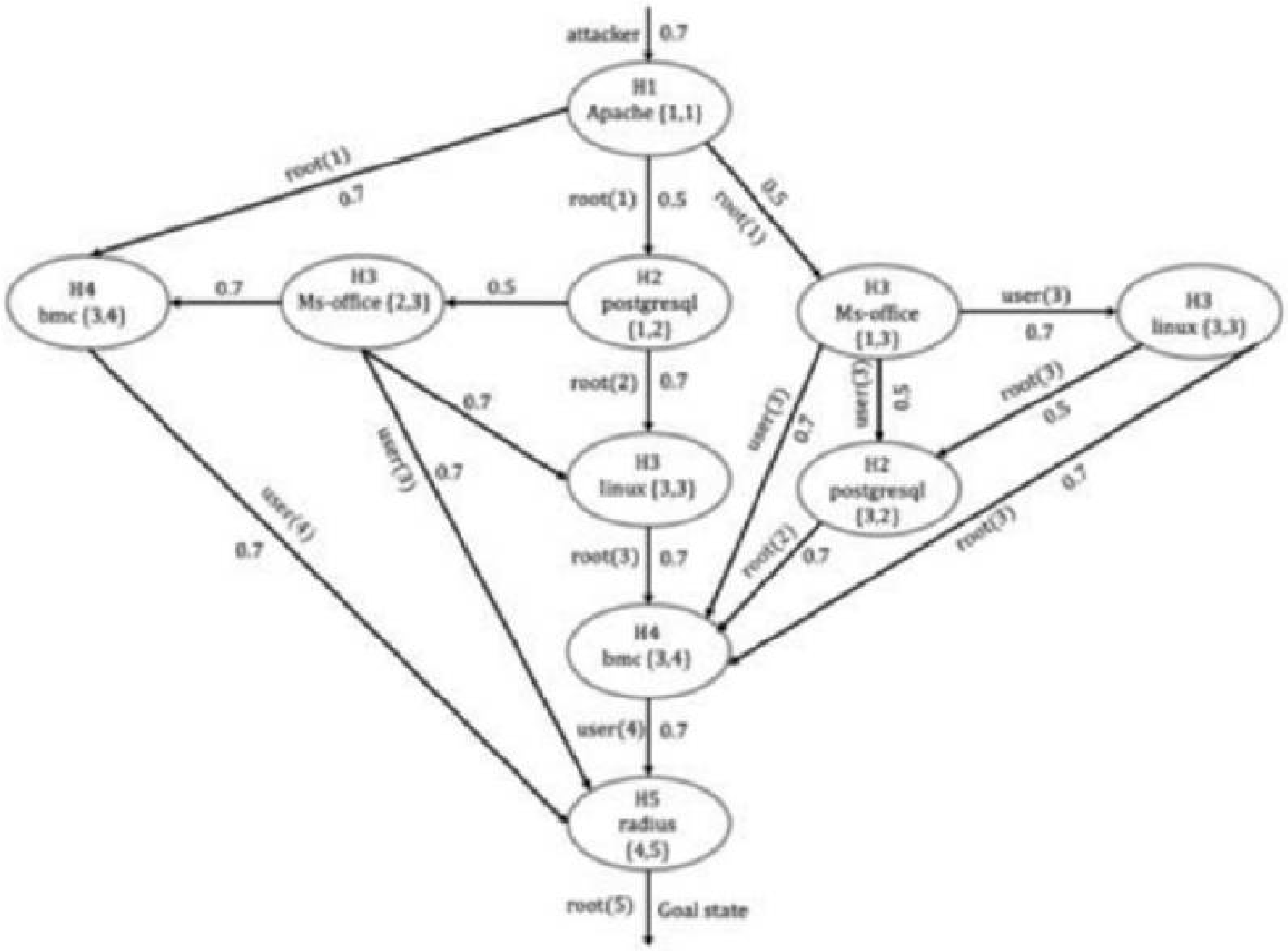


FIGURE 10. Attack graph of the experimental network.

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, #20.3.3

Thank You for all your Attention

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