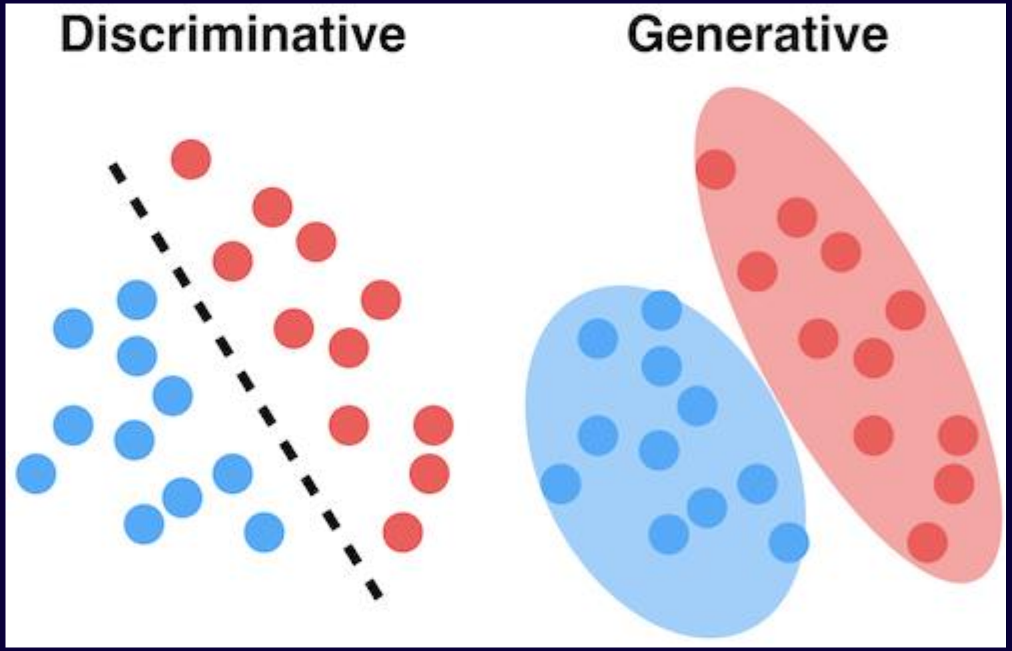
Hidden Markov Model

POS tagging using Generative Model

Types of Models

- 2 types
- Generative model
- Discriminative model



Motivation

- Variability is a part of nature
- Variability leads to uncertainty when drawing conclusion from data
- Motivates us to take a probablistic approach
- To modeling & reasoning

Motivation

- Probabilistic modelling:
  - Identify the quantities that relate to the aspects of reality that you wish to capture with your model.
  - Consider them to be random variables, e.g. $\mathbf{x}, \mathbf{y}, \mathbf{z}$, with a joint pdf (pmf) $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$.
- Probabilistic reasoning:
  - Assume you know that $\mathbf{y} \in \mathcal{E}$ (measurement, evidence)
  - Probabilistic reasoning about $\mathbf{x}$ then consists in computing

$$p(\mathbf{x}|\mathbf{y} \in \mathcal{E})$$

    or related quantities like its maximiser or posterior expectations.

Example

- 2 friends Rahul & Ashok
- Rahul - jog, go to the office, and cleaning his residence
- According to the weather he does the task
- He tells Ashok what he did
- Ashok assume weather as per the task
- Ashok believes weather {Rainy, Sunny} acts as a Markov chain

Example

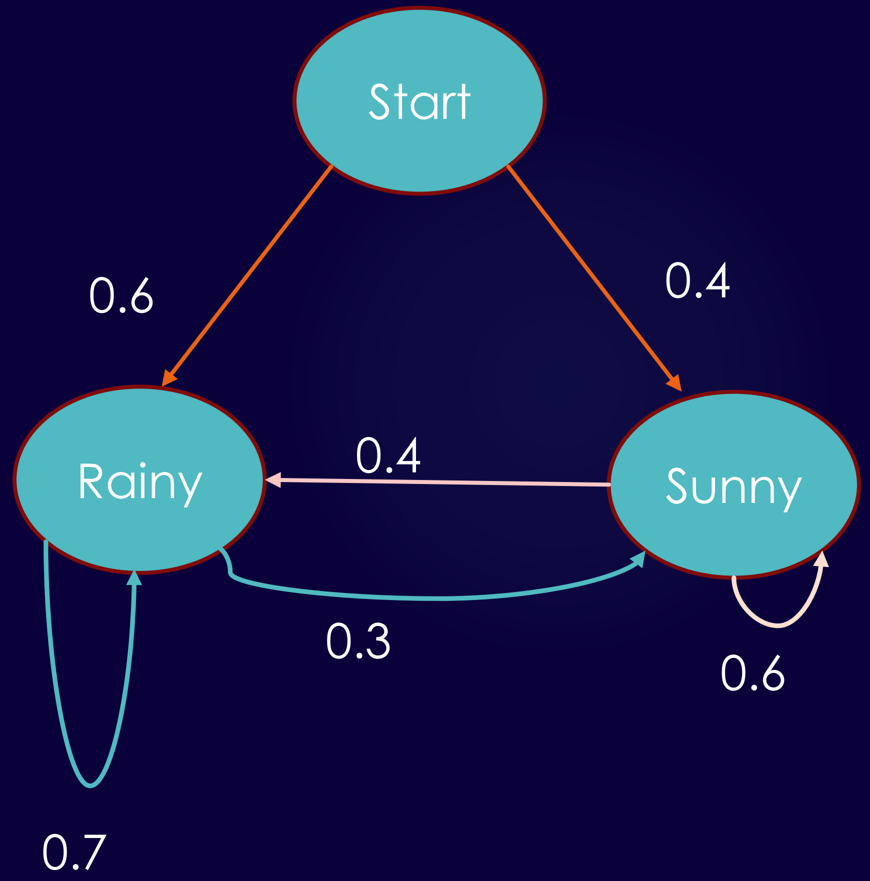
- On each day Rahul performs a task
- With certain probability
- Depending on weather
- Rahul tells Ashok he cleaned his home
- Ashok deducts from past experience that it was rainy day

Example

- The states and observation are:
- states = ('Rainy', 'Sunny')
- observations = (‘jog', ‘work', 'clean')
- And the start probability is:
- start_probability = {'Rainy': 0.6, 'Sunny': 0.4}

Example

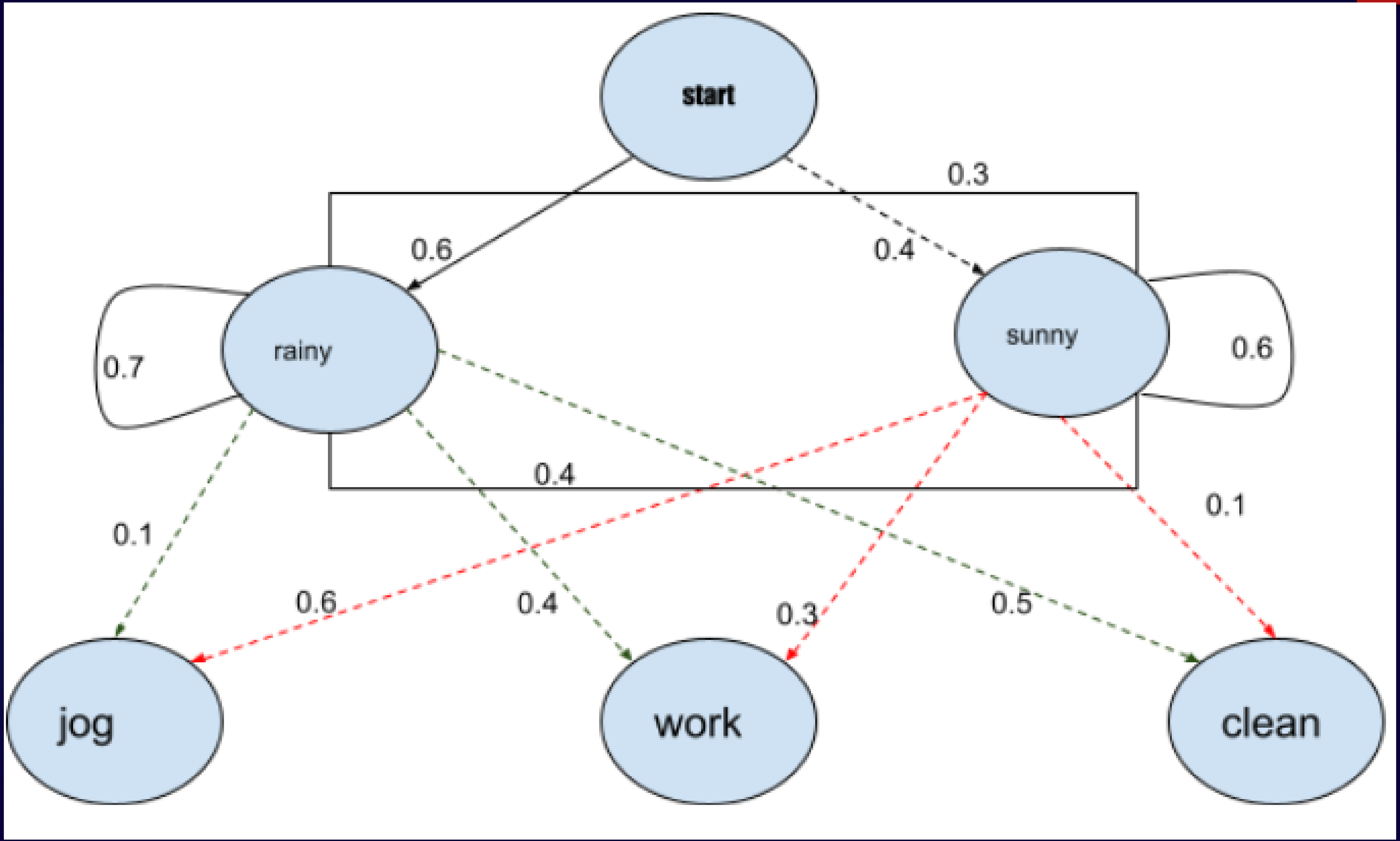
- transition_probability = {
- 'Rainy' : {'Rainy': 0.7, 'Sunny': 0.3},
- 'Sunny' : {'Rainy': 0.4, 'Sunny': 0.6},
- }



Example

- emission_probability = {
- 'Rainy' : {'jog': 0.1, 'work': 0.4, 'clean': 0.5},
- 'Sunny' : {'jog': 0.6, 'work: 0.3, 'clean': 0.1},
- }

Example



start
0.6
0.3
0.4
rainy
sunny
0.7
0.6
0.4
0.1
0.1
0.6
0.4
0.3
0.5
jog
work
clean

Applications of HMM

- Computational finance
- speed analysis
- Speech recognition
- Speech synthesis
- Part-of-speech tagging
- Document separation in scanning solutions
- Machine translation
- Handwriting recognition
- Time series analysis
- Activity recognition
- Sequence classification
- Transportation forecasting

Example 2

Problem

- Start W = w1,w2….wn words in the corpus (observed)
- T = t1…tn corresponding tags (unknown)

Tagging: Probabilistic View (Generative Model)

Find

$$\hat{T} = argmax_T P(T|W)$$

- Ie. Argmax P(t1…tn | w1…wn)

Using Bayes Theorem

- $W = w_1 \dots w_n$ - words in the corpus (observed)
- $T = t_1 \dots t_n$ - the corresponding tags (unknown)

Tagging: Probabilistic View (Generative Model)

Find

$$\begin{aligned}
\hat{T} &= argmax_T P(T|W) \\
&= argmax_T \frac{P(W|T)P(T)}{P(W)} \\
&= argmax_T P(W|T)P(T) \\
&= argmax_T \prod_i P(w_i|w_1 \dots w_{i-1}, t_1 \dots t_i)P(t_i|t_1 \dots t_{i-1})
\end{aligned}$$

Further simplification

$$\hat{T} = argmax_T \prod_i P(w_i|w_1 \ldots w_{i-1}, t_1 \ldots t_i) P(t_i|t_1 \ldots t_{i-1})$$

- The probability of a word appearing depends only on its own POS tag
  $P(w_i|w_1 \ldots w_{i-1}, t_1 \ldots t_i) \approx P(w_i|t_i)$
- Bigram assumption: the probability of a tag appearing depends only on the previous tag

Further simplification

$$\hat{T} = argmax_T \prod_i P(w_i|w_1 \ldots w_{i-1}, t_1 \ldots t_i) P(t_i|t_1 \ldots t_{i-1})$$

- The probability of a word appearing depends only on its own POS tag
  $P(w_i|w_1 \ldots w_{i-1}, t_1 \ldots t_i) \approx P(w_i|t_i)$
- Bigram assumption: the probability of a tag appearing depends only on the previous tag
  $P(t_i|t_1 \ldots t_{i-1}) \approx P(t_i|t_{i-1})$
- Using these simplifications:
  $\hat{T} = argmax_T \prod_i P(w_i|t_i) P(t_i|t_{i-1})$

What is this model?

Computing probability values

Tag Transition probabilities $p(t_i|t_{i-1})$

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

$$P(NN|DT) = \frac{C(DT, NN)}{C(DT)} = \frac{56,509}{116,454} = 0.49$$

Computing probability values

Tag Transition probabilities $p(t_i|t_{i-1})$

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

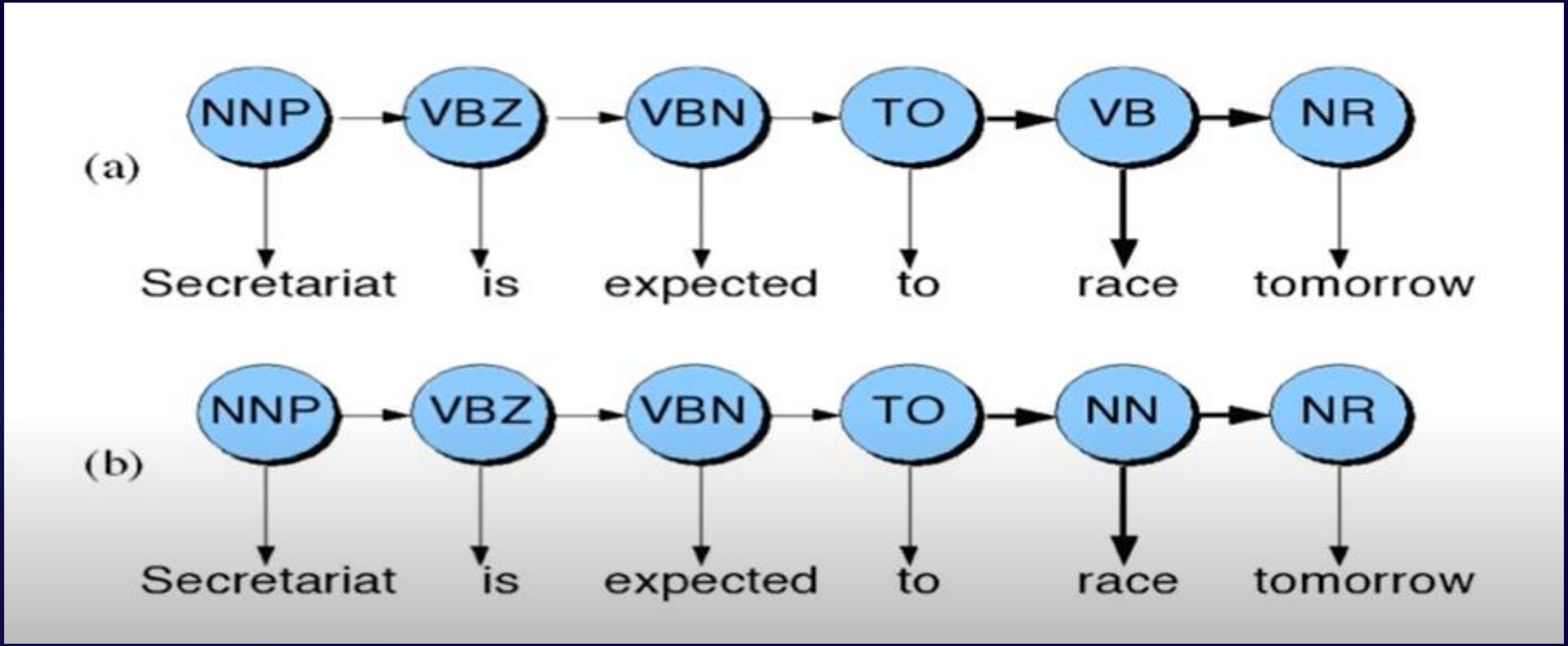
$$P(NN|DT) = \frac{C(DT, NN)}{C(DT)} = \frac{56,509}{116,454} = 0.49$$

Word Likelihood probabilities $p(w_i|t_i)$

$$P(w_i|t_i) = \frac{C(t_i, w_i)}{C(t_i)}$$

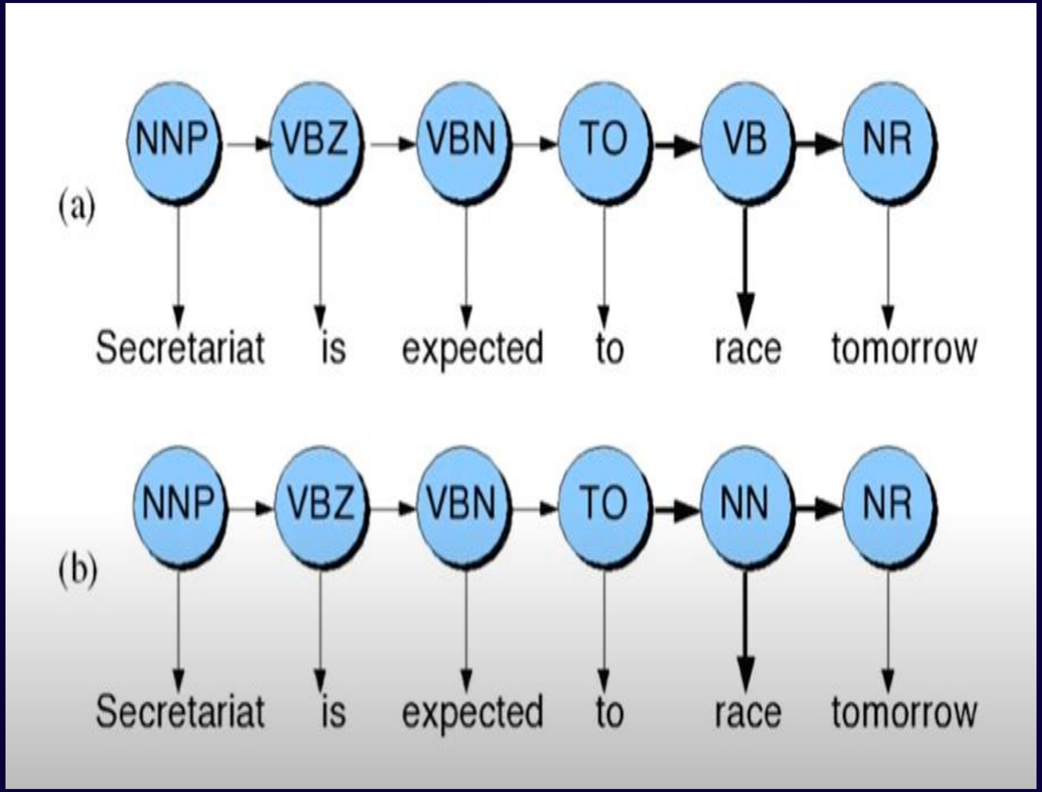
$$P(is|VBZ) = \frac{C(VBZ, is)}{C(VBZ)} = \frac{10,073}{21,627} = 0.47$$

Disambiguating "race"



Disambiguating “race”

- P(VB | TO) P(NN | TO)
- P(NR | VB) P(NR | NN)
- P(race | VB) P(race | NN)
- Mutiply the probabilities in each case
- Find higher probability



Disambiguating "race"

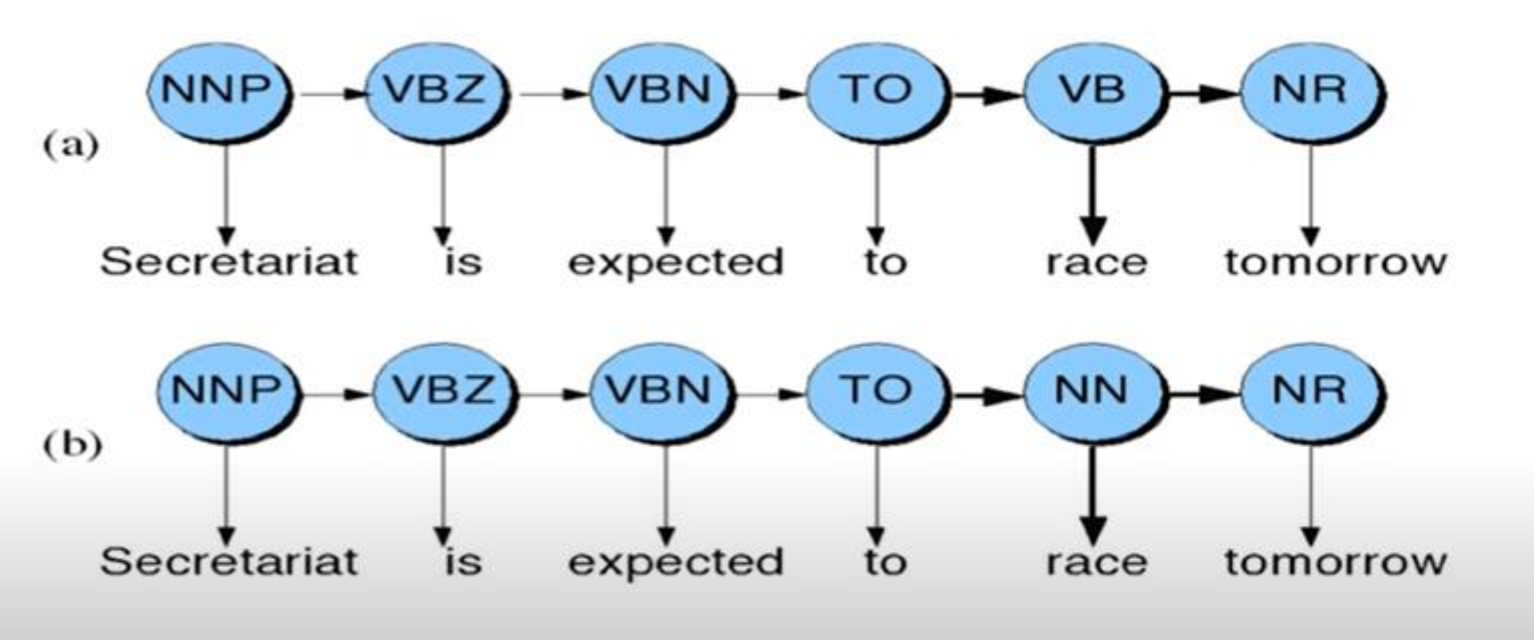
Difference in probability due to

- $P(VB|TO)$ vs. $P(NN|TO)$
- $P(race|VB)$ vs. $P(race|NN)$
- $P(NR|VB)$ vs. $P(NR|NN)$

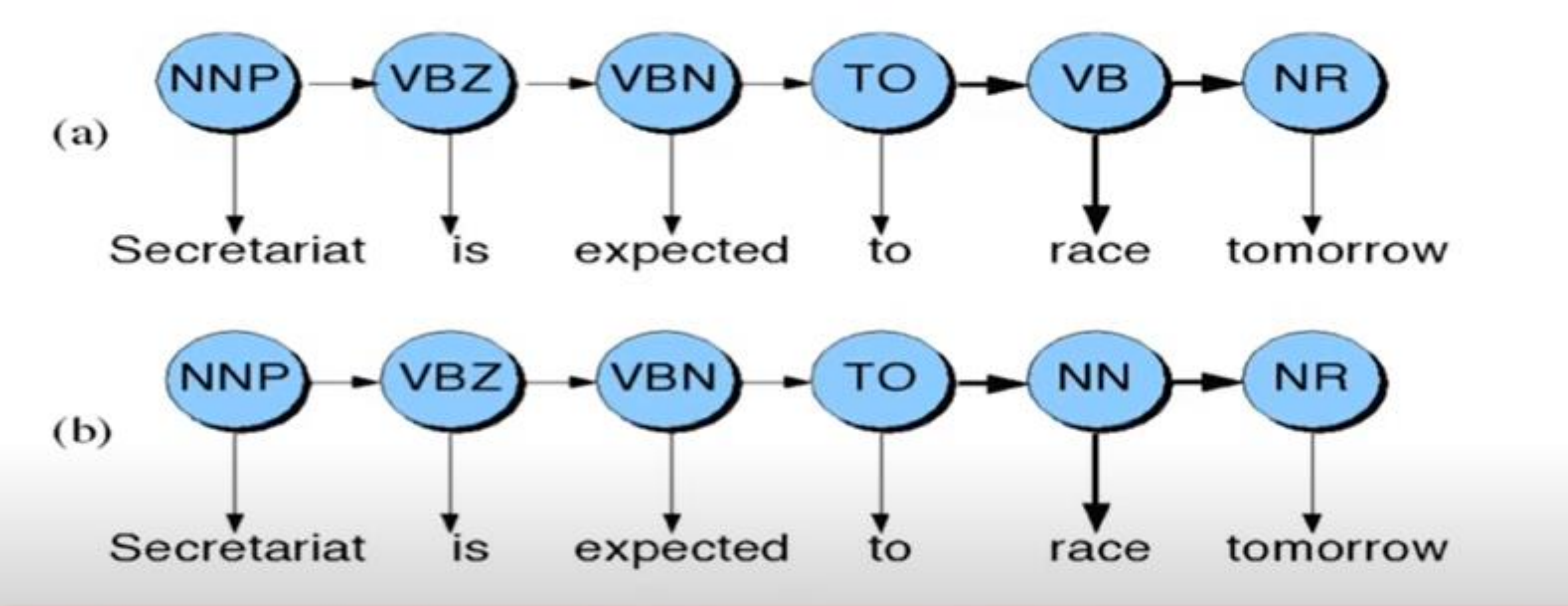
After computing the probabilities

- $P(NN|TO)P(NR|NN)P(race|NN) = 0.0047 \times 0.0012 \times 0.00057 = 0.00000000032$
- $P(VB|TO)P(NR|VB)P(race|VB) = 0.83 \times 0.0027 \times 0.00012 = 0.00000027$

What is this model?



What is this model?



This is a Hidden Markov Model

What is HMM?

- Tag transition probabilities $p(t_i \mid t_{i-1})$
- Word likelihood properties (emission) $p(w_i \mid t_i)$
- This is HMM

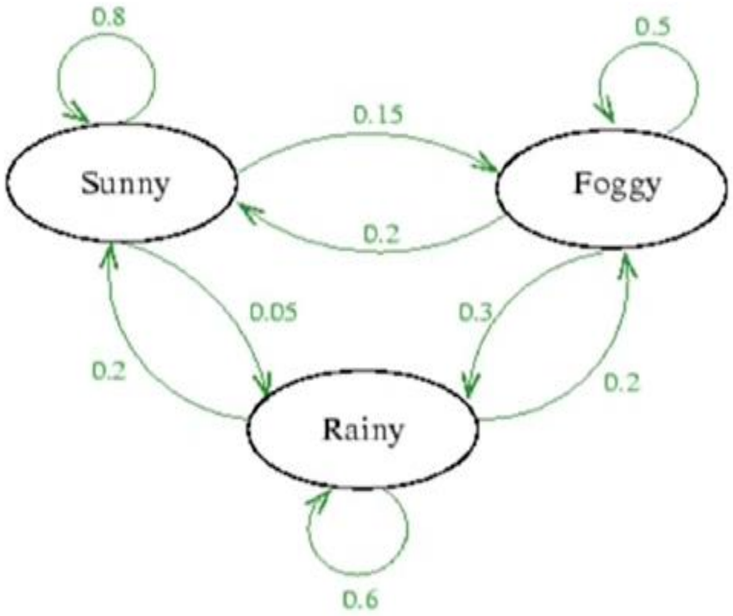
How is Markov model different from HMM

- Given today is sunny
- Probability of tomorrow sunny?
- Day after is rainy?
- $P(q_{n+2} = R, q_{n+1} = S)$
- $= P(q_{n+1} = S \mid q_n = S) *$
  $P(q_{n+2} = R \mid q_{n+1} = S, q_n = S)$
- Can be approximated as
- $= P(q_{n+1} = S \mid q_n = S) *$
  $P(q_{n+2} = R \mid q_{n+1} = S)$

$=0.8 * 0.05 = 0.04$

Table 1: Probabilities $p(q_{n+1}|q_n)$ of tomorrow's weather based on today's weather

	Tomorrow's weather		
Today's weather			
	0.8	0.05	0.15
	0.2	0.6	0.2
	0.2	0.3	0.5



How is Markov model different from HMM

- For Markov chains, the output symbols are the same as the states
  'sunny' weather is both observable and state
- But in POS tagging
  The output symbols are words
  But the hidden states are POS tags
- A Hidden Markov Model is an extension of a Markov chain in which the output symbols are not the same as the states
- We don't know which state we are in

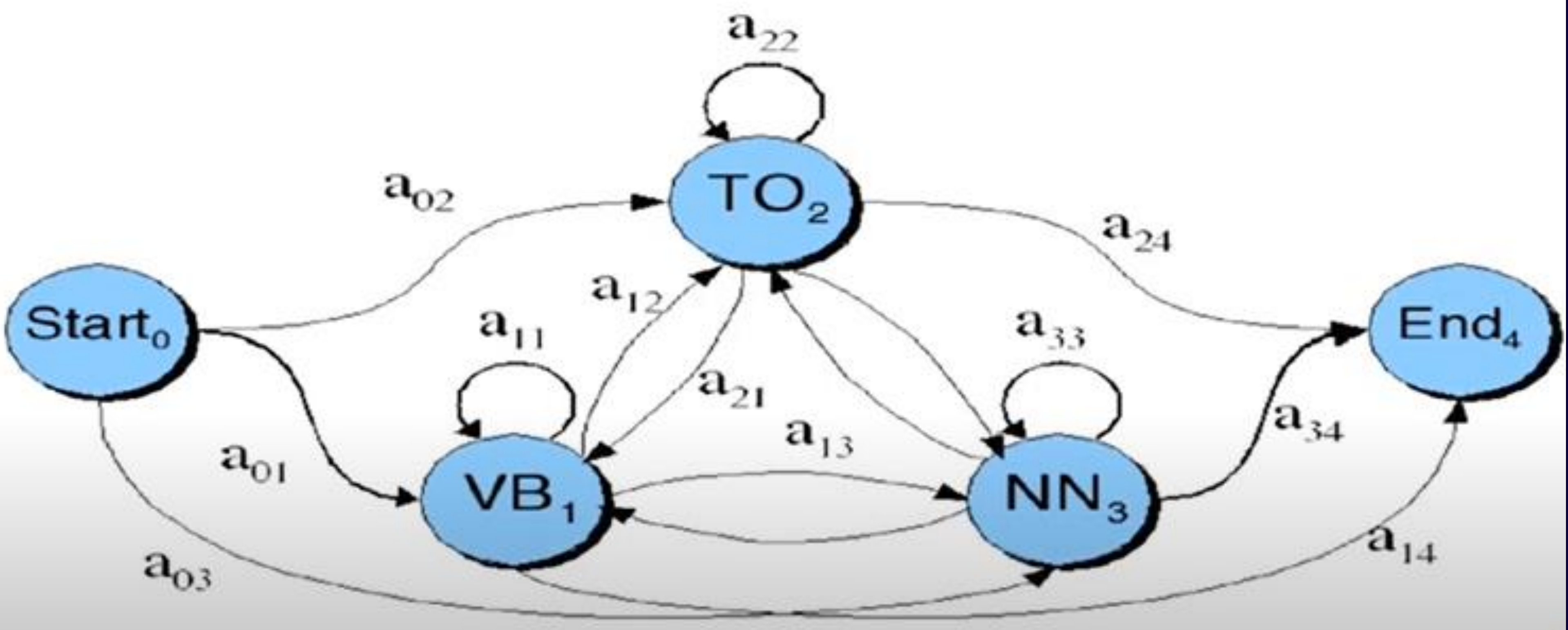
Elements of HMM model

Elements of an HMM model

- A set of states (here: the tags)
- An output alphabet (here: words)
- Initial state (here: beginning of sentence)
- State transition probabilities (here $p(t_n|t_{n-1})$)
- Symbol emission probabilities (here $p(w_i|t_i)$)

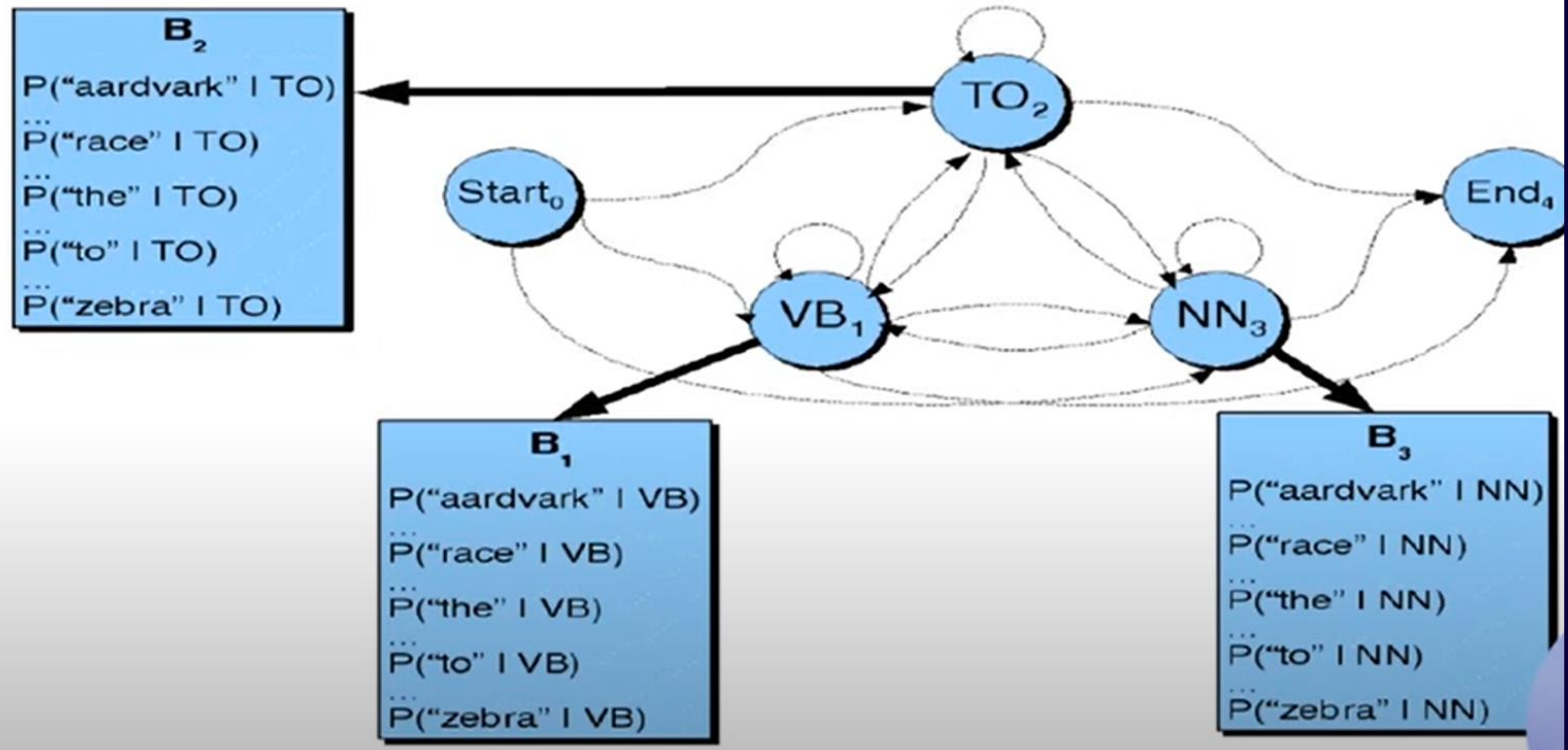
Graphical representation

When tagging a sentence, we are walking through the state graph:

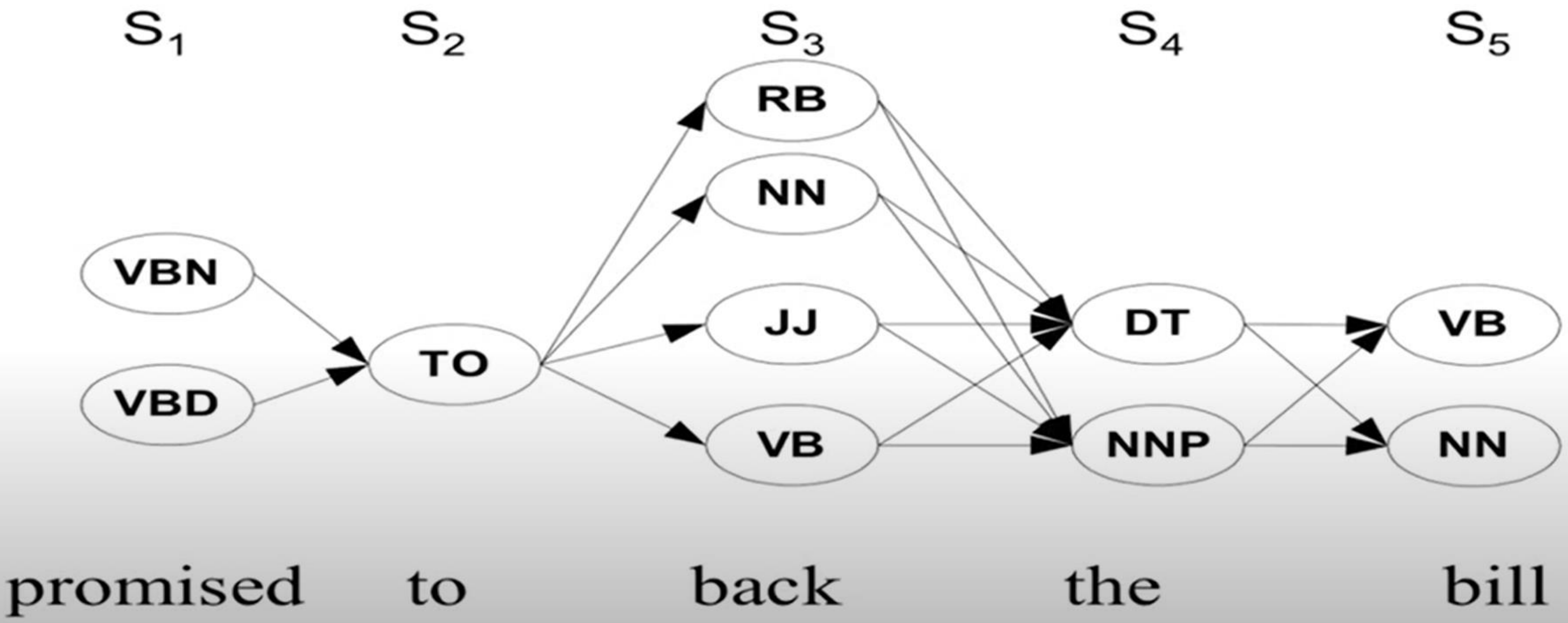


Graphical representation

At each state we emit a word: $P(w_n|t_n)$



Walking through the states to find the best path



Possible Approach

- Compute probability of each sequence
- This may not be efficient
- We need a solution that does not grow exponentially with the no. of words
- We will use VITERBI algorithm to do POS tag in efficient manner

Question

Question

- Calculate transition and emission probabilities for a set of sentences:

- Mary Jane can see will
- Spot will see Mary
- Will Jane spot Mary?
- Mary will pat Spot

Solution

Emission Probability

Words	Noun	Modal	Verb
mary			
jane			
will			
spot			
can			
see			
pat			

Transition Probabiliy

	Noun	Modal	Verb	End
Start				
Noun				
Model				
Verb				

Emission Probability

Words	Noun	Modal	Verb
mary	4	0	0
jane	2	0	0
will	1	3	0
spot	2	0	1
can	0	1	0
see	0	0	2
pat	0	0	1

Words	Noun	Modal	Verb
mary	4/9	0	0
jane	2/9	0	0
will	1/9	3/4	0
spot	2/9	0	1/4
can	0	1/4	0
see	0	0	2/4
pat	0	0	1/4

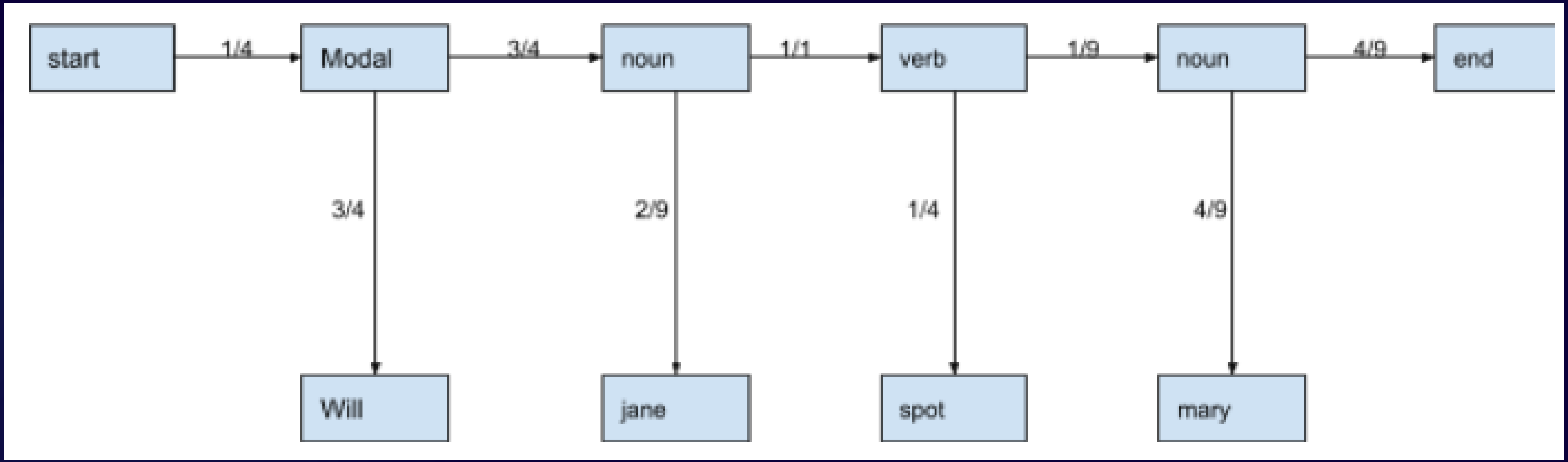
Transition Probabiliy

	Noun	Modal	Verb	End
Start	3	1	0	0
Noun	1	3	1	4
Model	1	0	3	0
Verb	4	0	0	0

Transition Probability

	Noun	Modal	Verb	End
Start	3/4	1/4	0	0
Noun	1/9	3/9	1/9	4/9
Model	1/4	0	3/4	0
Verb	4/4	0	0	0

Solution

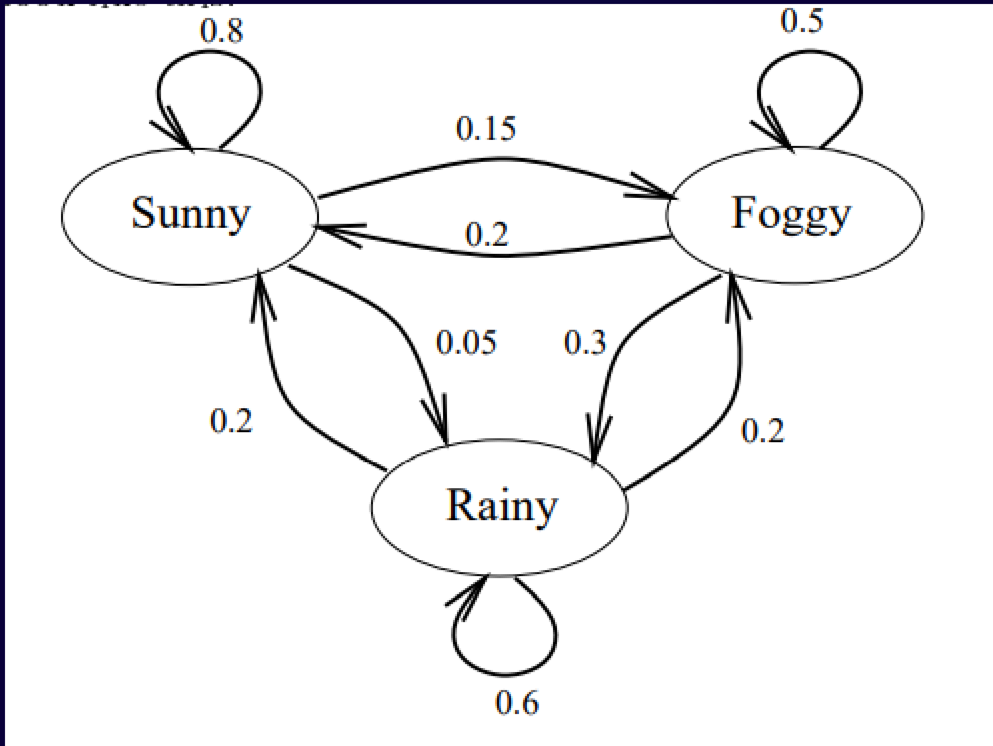


start
1/4
Modal
3/4
noun
1/1
verb
1/9
noun
4/9
end
3/4
2/9
1/4
4/9
Will
jane
spot
mary

Question

Question

- Due to COVID you were locked in a room for several days , and asked about the weather outside. The only piece of evidence is whether the person comes into the room carrying your daily meal is carrying an umbrella or not.
- Following probability is given.
- On a sunny day P(umbrella) = 0.1
- On a rainy day P(umbrella) = 0.8
- On a foggy day P(umbrella) = 0.3



Part -1

- Suppose the day you were locked in was sunny. Next day caretake carried an umbrella into the room. Assuming that prior probability of caretaker carrying an umbrella on any day is 0.5. what is the probability that second day was rainy?

Part -2

- Suppose the day you were locked in was sunny. The caretaker brought an umbrella on day 2 but not on day 3. Assuming prior probability of bringing an umbrella is 0.5 , what is the probability that it is foggy on day 3?

Part -1 Solution

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny}|u_2 = \text{T})} \\
(u_2 \ \ and \ \ w_1 \ \ independent) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

$$
\begin{aligned}
(Bayes' Rule) &= \frac{P(u_2 = \text{T}|w_1 = \text{Sunny}, w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Markov \ \ assumption) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})}
\end{aligned}
$$

$$
\begin{aligned}
(P(A, B) = P(A|B)P(B)) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}|w_1 = \text{Sunny})P(w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Cancel : P(Sunny)) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}|w_1 = \text{Sunny})}{P(u_2 = \text{T})} \\
&= \frac{(0.8)(0.05)}{0.5} \\
&= .08
\end{aligned}
$$

Part -2 Solution

$$
\begin{aligned}
P(w_3 = \mathrm{F} \mid \quad & = P(w_2 = \mathrm{Foggy}, w_3 = \mathrm{Foggy} \mid \\
w_1 = \mathrm{S}, u_2 = \mathrm{T}, u_3 = \mathrm{F}) \quad & \qquad w_1 = \mathrm{Sunny}, u_2 = \mathrm{True}, u_3 = \mathrm{False}) + \\
& \quad P(w_2 = \mathrm{Rainy}, w_3 = \mathrm{Foggy} \mid \ldots) + \\
& \quad P(w_2 = \mathrm{Sunny}, w_3 = \mathrm{Foggy} \mid \ldots) \\
& = \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \quad \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \quad \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)}
\end{aligned}
$$

Part -2 Solution

$$
\begin{aligned}
P(w_3 = \mathrm{F} \mid \qquad & = P(w_2 = \mathrm{Foggy}, w_3 = \mathrm{Foggy} \mid \\
w_1 = \mathrm{S}, u_2 = \mathrm{T}, u_3 = \mathrm{F}) \qquad & \qquad\quad w_1 = \mathrm{Sunny}, u_2 = \mathrm{True}, u_3 = \mathrm{False}) + \\
& P(w_2 = \mathrm{Rainy}, w_3 = \mathrm{Foggy} \mid \ldots) + \\
& P(w_2 = \mathrm{Sunny}, w_3 = \mathrm{Foggy} \mid \ldots) \\
& = \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)}
\end{aligned}
$$

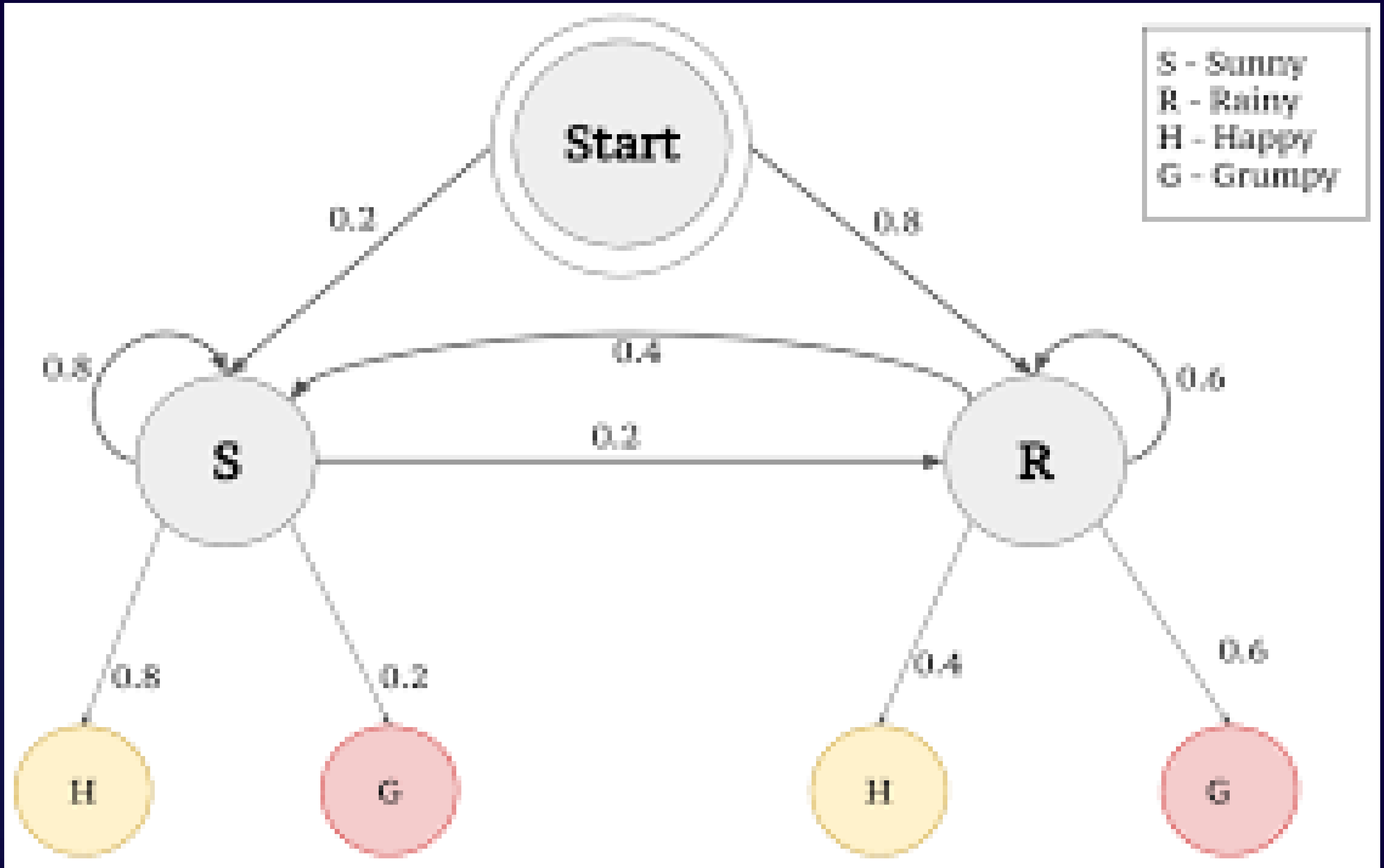
Part -2 Solution

$$= \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)}{P(u_3 = F)P(u_2 = T)} +$$
$$\frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)}{P(u_3 = F)P(u_2 = T)} +$$
$$\frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)}{P(u_3 = F)P(u_2 = T)}$$

Part -2 Solution

$$
\begin{aligned}
&= \frac{(0.7)(0.3)(0.5)(0.15)}{(0.5)(0.5)} + \\
&\quad \frac{(0.7)(0.8)(0.2)(0.05)}{(0.5)(0.5)} + \\
&\quad \frac{(0.7)(0.1)(0.15)(0.8)}{(0.5)(0.5)} \\
&= 0.119
\end{aligned}
$$

HMM Recap



S - Sunny
R - Rainy
H - Happy
G - Grumpy
Start
0.2
0.8
0.8
0.4
0.6
S
0.2
R
0.8
0.2
0.4
0.6
H
G
H
G