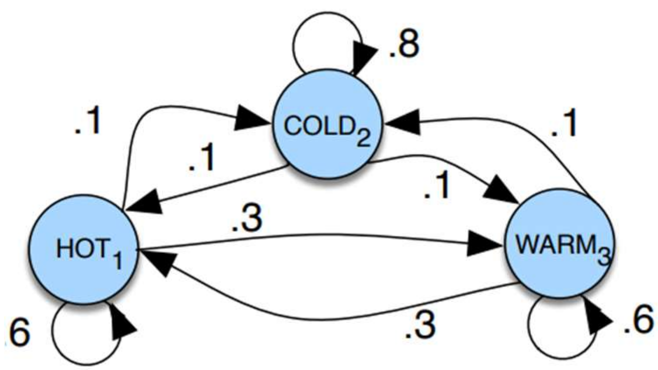
Hidden Markov Models

Dr Padmavathy T, SCOPE

Markov Chain

- A Markov chain is a model that tells us something about the probabilities of sequences of random variables, states, each of which can take on values from some set.
- These sets can be words, or tags, or symbols representing anything, like the weather.
- Future state is predicted based on the current state.



A start distribution π is required; setting π = [0.1, 0.7, 0.2] would mean,
A probability 0.7 of starting in state 2 (cold)
A probability 0.1 of starting in state 1 (hot), etc

Markov Chain

- Consider a sequence of state variables q1,q2,...,qi . A Markov model embodies the Markov assumption on the probabilities of this sequence: that when predicting the future, the past doesn't matter, only the present matters.

$$\textbf{Markov Assumption:} \quad P(q_i = a | q_1 ... q_{i-1}) = P(q_i = a | q_{i-1})$$

Hidden Markov Model

- A Markov chain is useful when we need to compute a probability for a sequence of observable events.
- In many cases, however, the events we are interested in are hidden.
- For example we don't normally observe part-of-speech tags in a text. Rather, we see words, and must infer the tags from the word sequence.
- A hidden Markov model (HMM) allows us to talk about both observed events (like words that we see in the input) and hidden events (like part-of-speech tags) that we think of as causal factors in our probabilistic model.
- Example – He runs a successful Business

Hidden Markov Model

- An HMM is specified by the following components:

$Q = q_1 q_2 \ldots q_N$	a set of N **states**
$A = a_{11} \ldots a_{ij} \ldots a_{NN}$	a **transition probability matrix** A, each a_{ij} representing the probability of moving from state i to state j, s.t. $\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$
$B = b_i(o_t)$	a sequence of **observation likelihoods**, also called **emission probabilities**, each expressing the probability of an observation o_t (drawn from a vocabulary $V = v_1, v_2, \ldots, v_V$) being generated from a state q_i
$\pi = \pi_1, \pi_2, \ldots, \pi_N$	an **initial probability distribution** over states. π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j = 0$, meaning that they cannot be initial states. Also, $\sum_{i=1}^{n} \pi_i = 1$

- First, as with a first-order Markov chain, the probability of a particular state depends only on the previous state:

$$\textbf{Markov Assumption:} \quad P(q_i|q_1...q_{i-1}) = P(q_i|q_{i-1})$$

- Second, the probability of an output observation oi depends only on the state that produced the observation qi and not on any other states or any other observations:

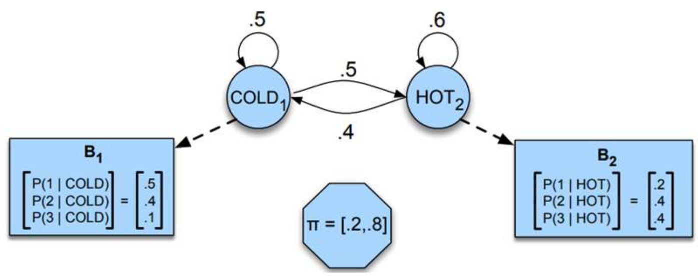
$$\textbf{Output Independence:} \quad P(o_i|q_1 \dots q_i, \dots, q_T, o_1, \dots, o_i, \dots, o_T) = P(o_i|q_i)$$

Hidden Markov Models

- Hidden Markov Model (HMM) is a statistical Markov model in which the model states are hidden and the state-dependent output of the model is visible.
- Applications of HMM
  - PoS Tagging
  - Speech recognition and Speech Synthesis
  - Gene Prediction
  - Machine Translation
  - Alignment of Bio Sequences

Hidden Markov Models

- Given a sequence of observations O (each an integer representing the number of ice creams eaten on a given day) find the 'hidden' sequence Q of weather states (H or C) which caused Jason to eat the ice cream.
- The two hidden states (H and C) correspond to hot and cold weather, and the observations (drawn from the alphabet O = {1,2,3}) correspond to the number of ice creams eaten by Jason on a given day.



Hidden Markov Models

- Hidden Markov models should be characterized by three fundamental problems:

Problem 1 (Likelihood): Given an HMM $\lambda = (A, B)$ and an observation sequence O, determine the likelihood $P(O|\lambda)$.

Problem 2 (Decoding): Given an observation sequence O and an HMM $\lambda = (A, B)$, discover the best hidden state sequence Q.

Problem 3 (Learning): Given an observation sequence O and the set of states in the HMM, learn the HMM parameters A and B.

Hidden Markov Models

- Let us consider our friends **Mohan** come form a distant city.
- Based on Weather condition in his city, he either carries an Umbrella **U** or doesn't carry an Umbrella **NU**.
- we can observe only whether he is carrying an umbrella or not; but not the weather condition of his city.
- Further assume that Mohan city has one of three possible weather conditions
  - Sunny
  - Cloudy
  - Runny

Hidden Markov Models

- HMM model consists of Five tuple - $\{\mathbf{Q}, \mathbf{O}, \mathbf{A}, \mathbf{B}, \pi\}$
  - set of States Q :
    - **N=3 states**
    - $\{Sunny(q_1), Cloudy(q_2), Rainy(q_3)\}$
  - set of observations O:
    - **M=2 no of unique observations**
    - $\{U(Umbrella), NU(NotUmbrella)\}$
  - state transition matrix (A) -
  - emission matrix (B)
  - Initial state probability (π)
  - Model parameter: $\lambda = < A, B, \pi >$

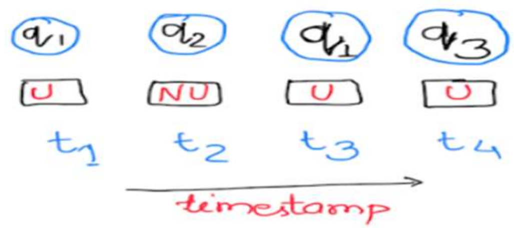
Hidden Markov Models

$$\mathbf{A} = \begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \end{array} \begin{array}{c} \begin{array}{ccc} q_1 & q_2 & q_3 \end{array} \\ \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.1 & 0.2 & 0.7 \\ 0.4 & 0.25 & 0.35 \end{pmatrix} \end{array}$$

$$\mathbf{B} = \begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \end{array} \begin{array}{c} \begin{array}{cc} U & NU \end{array} \\ \begin{pmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \\ 0.9 & 0.1 \end{pmatrix} \end{array}$$

$$\pi = \begin{array}{c} \begin{array}{ccc} q_1 & q_2 & q_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \end{pmatrix} \end{array}$$

Q1: **Calculate the joint probability of following sequence of observation and states**: P(<U NU U U>, $<q_1, q_2, q_1, q_3>$)



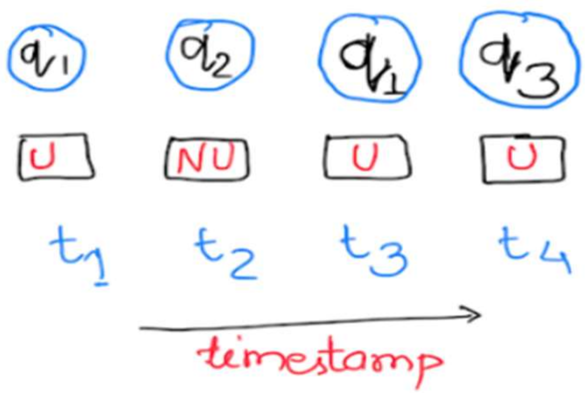
Calculation:

1. $\pi[q_1]$* $\mathbf{B}[U|q_1] = 0.5 * 0.3 = 0.15$
2. $\mathbf{A}([q_1][q_2])$ * $\mathbf{B}[NU|q_2] = 0.3 * 0.8 = 0.24$
3. $\mathbf{A}([q_2][q_1])$ * $\mathbf{B}[U|q_1] = 0.1 * 0.3 = 0.03$
4. $\mathbf{A}([q_1][q_3])$ * $\mathbf{B}[U|q_3] = 0.5 * 0.9 = 0.45$

Hidden Markov Models

Q1:**Calculate the joint probability of following sequence of observation and states**: P(<U NU U U>, <q_1, q_2, q_1, q_3 >)

Calculation:



1. $\pi[q_1]$* $\mathbf{B}[\mathrm{U}|q_1] = 0.5$ *0.3 = 0.15
2. $\mathbf{A}([q_1][q_2])$ * $\mathbf{B}[\mathrm{NU}|q_2] = 0.3$ *0.8 =0.24
3. $\mathbf{A}([q_2][q_1])$ * $\mathbf{B}[\mathrm{U}|q_1] = 0.1$ *0.3 =0.03
4. $\mathbf{A}([q_1][q_3])$ * $\mathbf{B}[\mathrm{U}|q_3] = 0.5$ *0.9 =0.45
5. final prob = 0.15 * 0.24 * 0.03 * 0.45
   = 0.000486

Hidden Markov Models – Likelihood Computation

- Computing Likelihood: Given an HMM λ = (A,B) and an observation sequence O, determine the likelihood $P(O|\lambda)$
- Note :
  - In Hidden Markov models, each hidden state produces only a single observation.
  - Thus, the sequence of hidden states and the sequence of observations have the same length.
- Algorithms
  - Forward Algorithm
  - Backward Algorithm

Example : Find the probability of the sequence U, NU, U, U from above example.

Likelihood Computation – Forward Algorithm

Given Model parameter $\lambda =< A, B, \pi >$ and a sequence of observation $o_1, o_2, o_3, ..., o_{T-1}, o_T$

We define a variable (helper function) α_t^j

$$\alpha_{\mathbf{t}}^{\mathbf{j}} = P(O_1, O_2, O_3, ..., O_t | S_t = q_j, \lambda)$$

- We have calculated probability of the sequence $o_1, o_2, o_3, ..., o_{t-1}, o_t$
- the last sequence is emitted from the state q_j.

Hidden Markov Models – Forward Algorithm

1. **Initialization** (Base case):

   For $j \in [1, N]$:

   $$\alpha_1^j = \pi_j * b_j(o_1)$$

2. **Inductive case**:

   For $t \in [2, T]$:

   For $j \in [1, N]$:

   $$\alpha_t^j = \sum_{i=1}^{N} \alpha_{t-1}^i * a_{ij} * b_j(o_t) \; ;$$

3. **Termination**:

   $$P(\mathbf{O}|\lambda) = \sum_{j=1}^{j=N} \alpha_T^j$$

What is the probability of the observed sequence " SSH" given the parameters of the HMM?

$$A = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

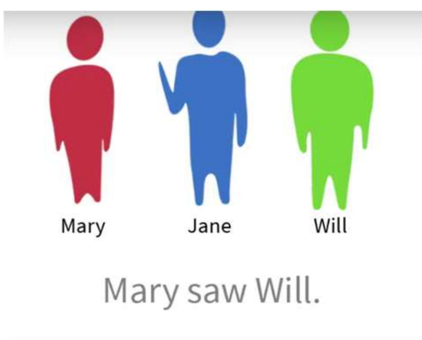
$$\pi = [0.375 \quad 0.625]$$

Hidden Markov Models – PoS Tagging

- Some applications - > Grammatical Analyzer (Spell Checkers)

Hidden Markov Models – PoS Tagging

- Noun - John, Car, India , Apple
- Verb – Run, Swim, Speak
- Modal Verb – Must, Will, Would, May
- Be verb – am, is, are, was, were, been, being



Jane	saw	Will	Marry	saw	Jane
noun	verb	noun	noun	verb	noun

Hidden Markov Models – PoS Tagging

	N	V
Mary	1	0
saw	0	2
Jane	2	0
Will	1	0

Hidden Markov Models – PoS Tagging

Marry	will	see	will
noun	*modal*	*verb*	*???*

Our data!

Mary	**will**	**see**	**Jane**
noun	*modal*	*verb*	*noun*
Will	**will**	**see**	**Mary**
noun	*modal*	*verb*	*noun*
Jane	**will**	**see**	**Will**
noun	*modal*	*verb*	*noun*

	N	V	M
Mary	2	0	0
see	0	3	0
Jane	2	0	0
Will	2	0	3

Hidden Markov Models – PoS Tagging

Marry	will	see	will
noun	*modal*	*verb*	*noun*

Our data!

Mary	**will**	**see**	**Jane**
noun	*modal*	*verb*	*noun*
Will	**will**	**see**	**Mary**
noun	*modal*	*verb*	*noun*
Jane	**will**	**see**	**Will**
noun	*modal*	*verb*	*noun*

BIGRAMS

	N-M	M-V	V-N
mary-will	1	0	0
will-see	0	3	0
see-jane	0	0	1
will-will	1	0	0
see-mary	0	0	1
jane-will	1	0	0
see-will	0	0	1

Hidden Markov Models – PoS Tagging

- Emission Probabilities – How likely is that Jane will be a noun, will be a modal , ….
- Transition Probabilities – How likely is it that ….. Noun is followed by a modal, which is followed by a verb,…….

Hidden Markov Models – PoS Tagging

Emission Probabilities

	N	M	V
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	1/2
Pat	0	0	1/4

N N M V N
Mary Jane can see Will.

N M V N
Spot will see Mary.

M N V N
Will Jane spot Mary?

N M V N
Mary will pat Spot

Transition Probabilities

	N	M	V	<E>
<S>	3	1	0	0
N	1	3	1	4
M	1	0	3	0
V	4	0	0	0

<S> N N M V N <E>
Mary Jane can see Will.

<S> N M V N <E>
Spot will see Mary.

<S> M N V N <E>
Will Jane spot Mary?

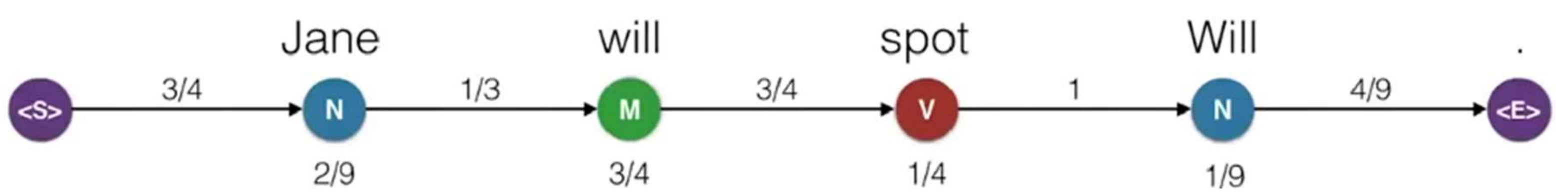
<S> N M V N <E>
Mary will pat Spot

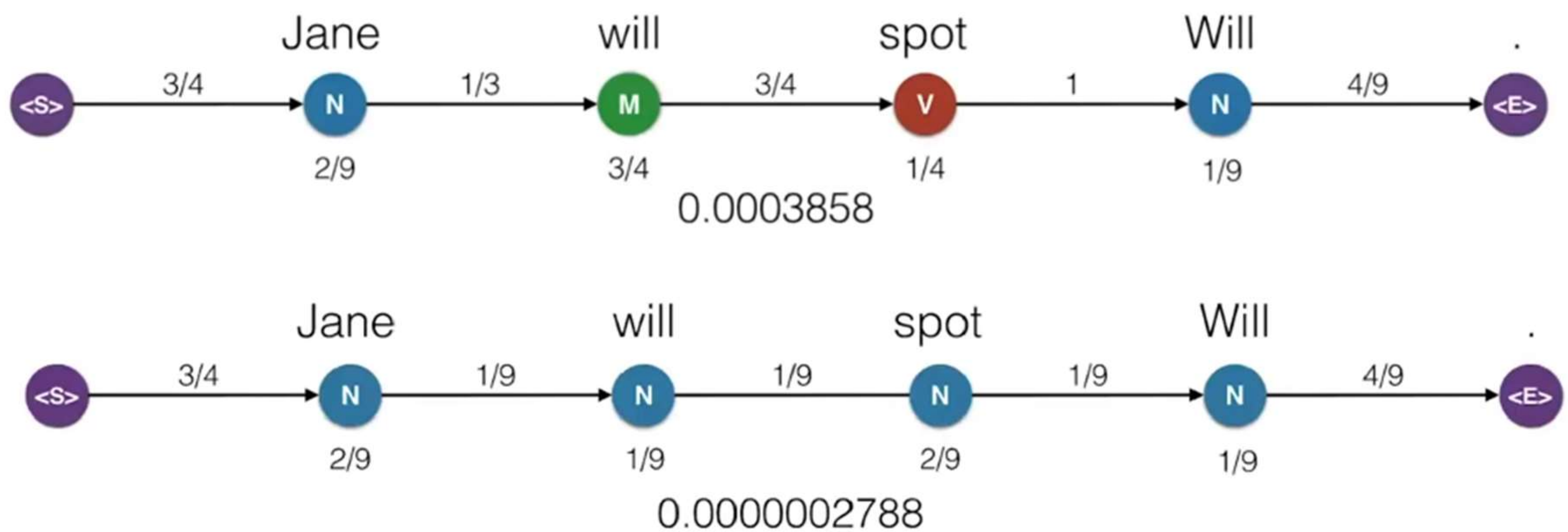
	N	M	V	<E>
<S>	3/4	1/4	0	0
N	1/9	1/3	1/9	4/9
M	1/4	0	3/4	0
V	1	0	0	0

Hidden Markov Models – PoS Tagging

- What is the probability that the following sentence occurs?

Jane will spot Will




Jane
will
spot
Will
.
<S>
3/4
N
1/3
M
3/4
V
1
N
4/9
<E>
2/9
3/4
1/4
1/9
0.0003858
Jane
will
spot
Will
.
<S>
3/4
N
1/9
N
1/9
N
1/9
N
4/9
<E>
2/9
1/9
2/9
1/9
0.0000002788

- How many different possibilities ? 3 ^4 = 81

Challenge ?

Determine the initial probability distribution $\pi = \pi_1, \pi_2, ..., \pi_N$ over the states ?

To Do……

- Use the following paragraph to estimate the state transition probability (A), initial state probability (π), and output emission probability (B) of POS tag and observed words using Hidden Markov Model.
- Did/VERB Brown/NOUN drink/VERB coffee/OTHER
- Coffee/NOUN is/VERB brown/OTHER
- Swift/NOUN is/VERB a/OTHER car/NOUN
- A/OTHER car/NOUN is/VERB brown/NOUN
- Brown/NOUN had/VERB a/Other swift/Other drive/NOUN
- Did/VERB Brown/VERB drive/VERB

Calculate the probability of the observation below using the Forward algorithm of the Hidden Markov Model.

A brown swift

References

- Daniel Jurafsky and James H Martin “Speech and Language Processing”, Prentice Hall 2017.
- Nitin Indurkhya, Fred J Damerau , “ Handbook of Natural language Processing”
- Part-of-speech (POS) tagging with Hidden Markov Model (HMM) (youtube.com)