ECE365: Introduction to NLP

Spring 2021

Lecture 5: Words in a sequence – Sequence labeling

[Reading J&M Chapter 8 (up to and including 8.4)]

Logistics

- Lab 3 is up

Course Progress

INTRODUCTION → WORD-LEVEL → WORD SEQUENCE → WORD MEANING → APPLICATION

WEEK 3

What is the nature of understanding we can get considering words as sequences?

- Mr. Forever who lives dangerously thinks he has nine lives.

- We ate in the afternoon and went on to have an afternoon tea.

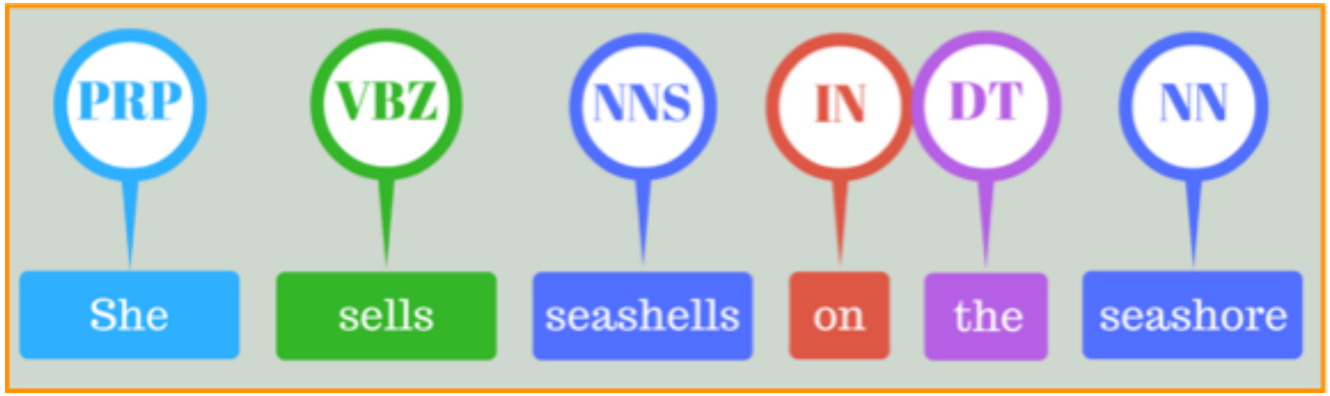
What is sequence labeling?

- Input: a sequence of word tokens **w**
- Output: a sequence of tags **t**, one per word (t$\in \mathcal{T}$)

What is sequence labeling?

Pronoun	Verb	Noun	Preposition	Determiner	Noun
She	sells	seashells	on	the	seashore

Sequence labeling tasks

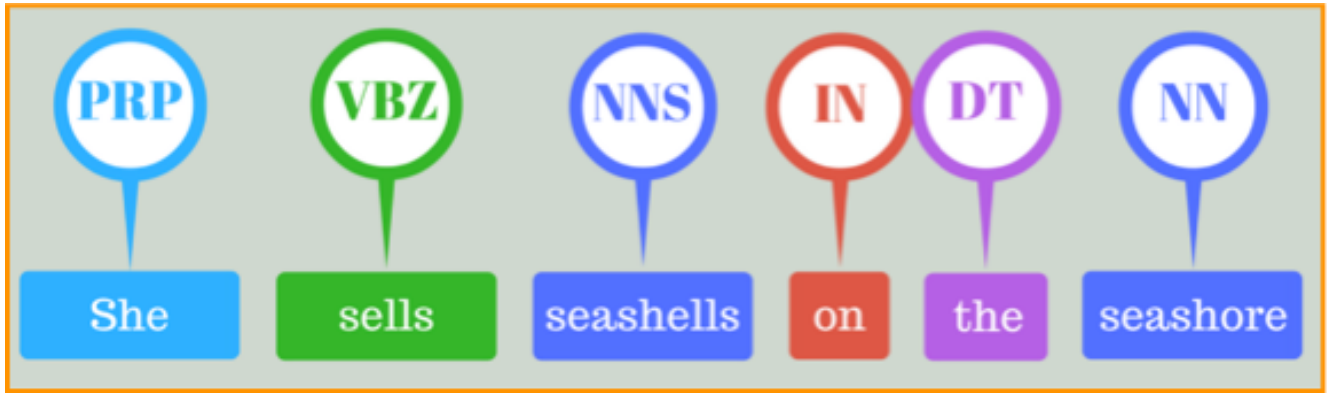


Part of Speech (POS) Tagging

Named Entity Recognition (NER)

Overview

POS tagging



Hidden Markov Model (HMM)

Viterbi Algorithm

What is sequence labeling?

- Input: a sequence of word tokens **w**

- Output: a sequence of POS tags **t**, one per word (t$\in \mathcal{T}$)

Part Of Speech Tagging

What are POS tags?

- Word classes
  - Nouns, Verbs, Adjectives, Adverbs
  - Prepositions, Conjunctions, Auxiliary verbs, Pronouns, determiners, numerals

POS tags

Open Classes

- Nouns
- Verbs
- Adjectives
- Adverbs

Closed Classes

- Pronouns
- Determiners
- Auxiliary verbs
- Prepositions
- Conjunctions
- Particles
- numerals

Google Universal POS Tags

- **ADJ**: adjective
- **ADP**: adposition (preposition or postposition)
- **ADV**: adverb
- **AUX**: auxiliary
- **CCONJ**: coordinating conjunction
- **DET**: determiner
- **INTJ**: interjection
- **NOUN**: noun
- **NUM**: numeral
- **PART**: particle
- **PRON**: pronoun
- **PROPN**: proper noun
- **PUNCT**: punctuation
- **SCONJ**: subordinating conjunction
- **SYM**: symbol
- **VERB**: verb
- **X**: other

Why do POS tagging?

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

Pronoun verb adjective noun! pronoun verb adjective, conjunction preposition adjective noun. Determiner….

POS tagging permits abstraction, allowing models to be more general

Why do POS tagging?

Text-to-Speech (how to pronounce the following words?)

English

- Transport
- Object. (She did not object to taking the object with her.)
- Discount
- Address
- Content

French: est, president

Useful for machine translation

How do humans assign tags?

- Jabberwocky (by Lewis Carroll 1872)

‘Twas brillig, and the slithy toves

Did gyre and gimble in the wabe:

All mimsy were the borogoves,

And the mome raths outgrabe.

Why is POS tagging hard?

earnings growth took a **back/JJ** seat
a small building in the **back/NN**
a clear majority of senators **back/VBP** the bill
Dave began to **back/VB** toward the door
enable the country to buy **back/RP** about debt
I was twenty-one **back/RB** then

Tag ambiguity: each word may have multiple POS tags

11% of all word types or 40% of all word tokens in Brown corpus (1M words) are ambiguous

What resources are available?

Some PTB Data (POS Tags)

IN In DT an NNP Oct. CD 19 NN review IN of `` `` DT The NN Misanthrope '' '' IN at NNP Chicago POS 's NNP Goodman NNP Theatre -LRB- -LRB- `` `` VBN Revitalized NNS Classics

VBP Take DT the NN Stage IN in NNP Windy NNP City , , '' '' NN Leisure CC & NNS Arts -RRB- -RRB- , , DT the NN role IN of NNP Celimene , , VBN played IN by NNP Kim NNP Cattrall , , VBD was RB mistakenly VBN attributed TO to NNP Christina NNP Haag . .

NNP Ms. NNP Haag VBZ plays NNP Elianti . .

NNP Rolls-Royce NNP Motor NNPS Cars NNP Inc. VBD said PRP it VBZ expects PRP$ its NNP U.S. NNS sales TO to VB remain JJ steady IN at IN about CD 1,200 NNS cars IN in CD 1990 . .

DT The NN luxury NN auto NN maker JJ last NN year VBD sold CD 1,214 NNS cars IN in DT the NNP U.S.

Is POS tagging a solved problem?

- **Most frequent class:** Assign each word token to the tag with which it occurred most in the training set. (e.g. back/NN) gives 90% accuracy

- **State of the art:** 97% accuracy at word level

- Average English sentence ~ 14 words

  Sentence level accuracies: 0.97^14 = **65%**

POS tagging not solved yet!

Techniques of POS tagging

- Rule based approaches
- Machine-learning methods – Hidden Markov Model

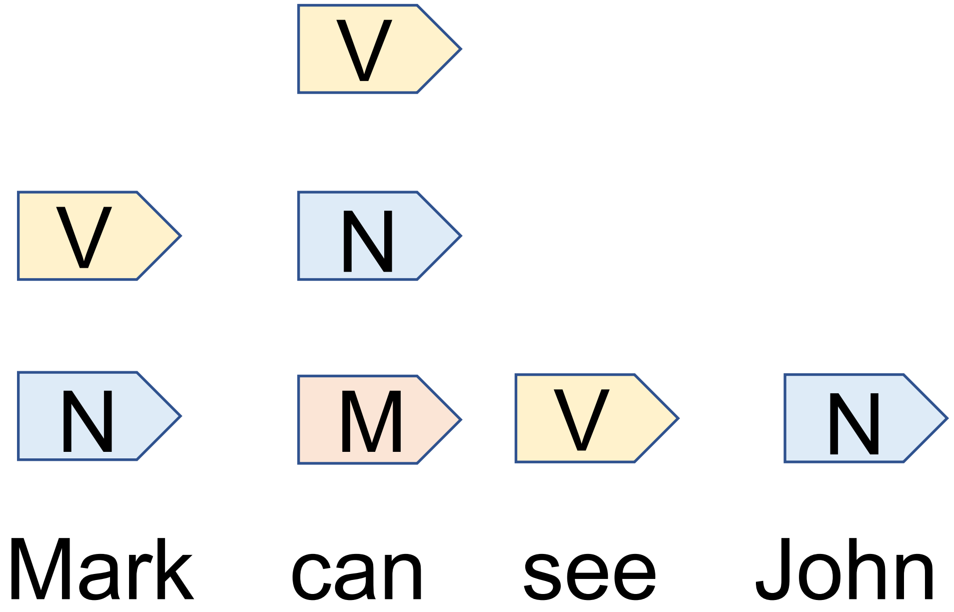
Noun (N)

Verb (V)

Modal verb (M)

special class of auxiliary verbs to express conditionality, necessity, obligation, ability, and wishful desire

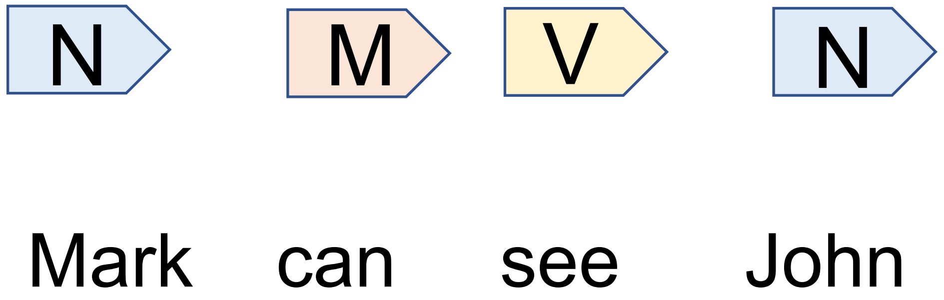
{shall, should, will, would, may, might, can, could, must}



Taking context into account can help

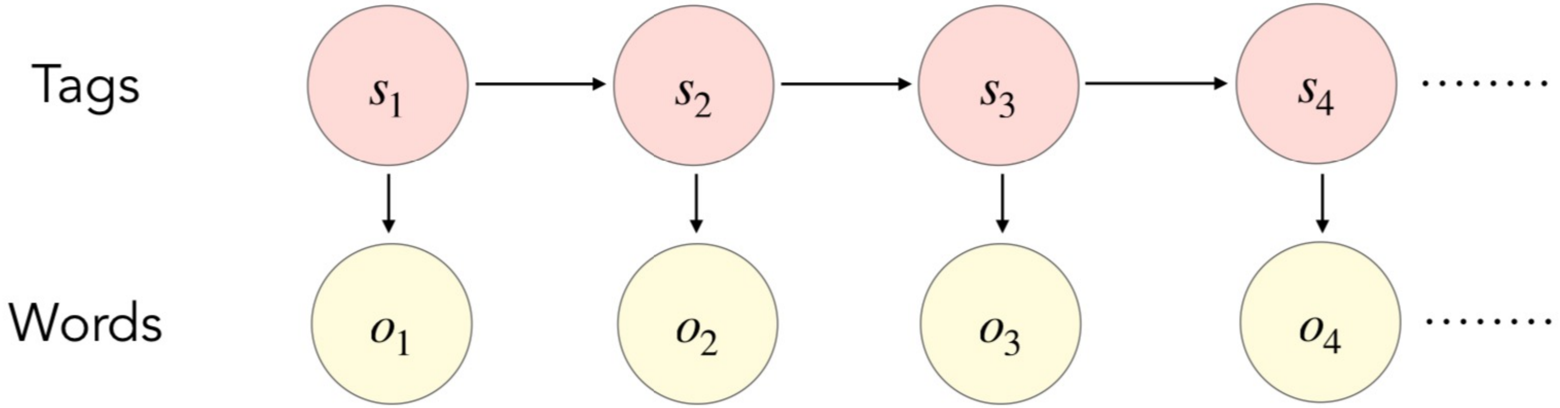
Taking context into account can help

Hidden Markov Model



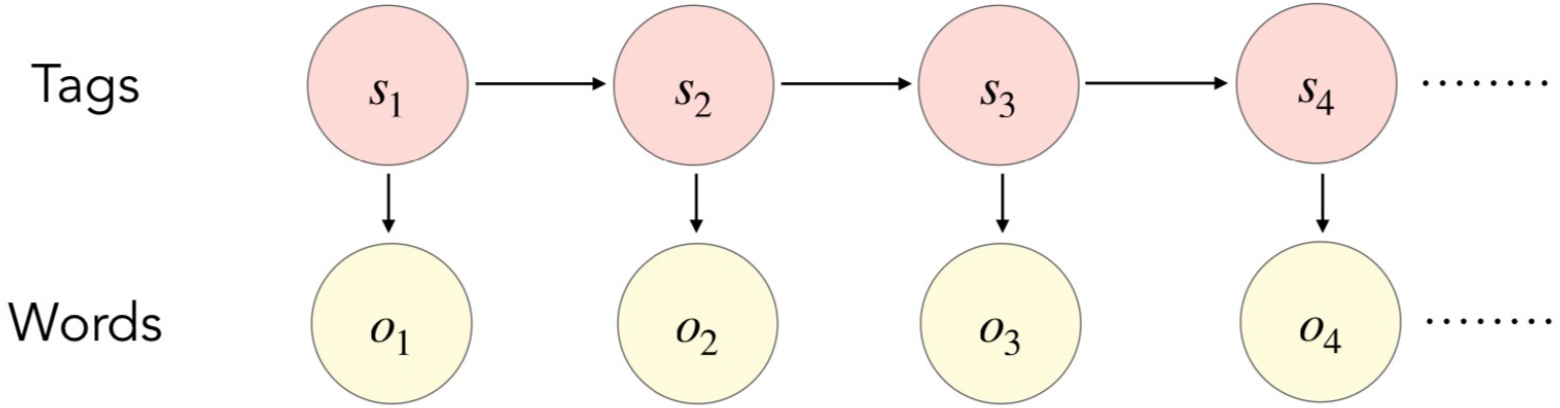
- Transition probabilities
- Emission probabilities
- Initial probabilities over tags

Hidden Markov Model



1. Set of states S = {1, 2, ..., K} and set of observations O
2. Initial state probability distribution $\pi(s_1)$
3. Transition probabilities $P(s_{t+1} \mid s_t)$
4. Emission probabilities $P(o_t \mid s_t)$

Hidden Markov Model Assumptions



1. Markov assumption:

$$P(s_{t+1} \mid s_1, \ldots, s_t) \approx P(s_{t+1} \mid s_t)$$

2. Output independence:

$$P(o_t \mid s_1, \ldots, s_t) \approx P(o_t \mid s_t)$$

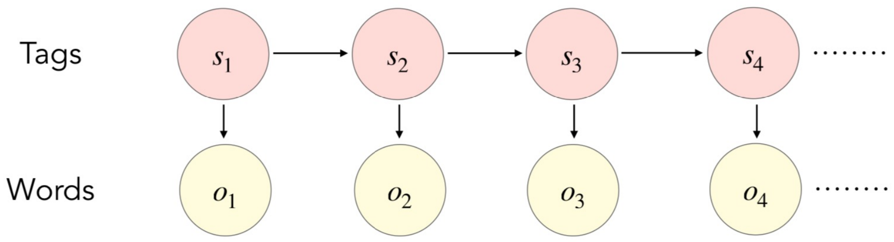
Sequence Likelihood

- What is the probability of

Mark/N can/M see/V John/N

- Joint probability

P(N, M, V, N, Mark, can, see, John)



1. Markov assumption:

$$P(s_{t+1} \mid s_1, \ldots, s_t) \approx P(s_{t+1} \mid s_t)$$

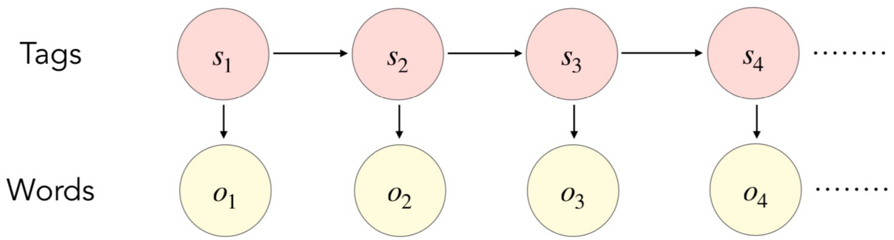
2. Output independence:

$$P(o_t \mid s_1, \ldots, s_t) \approx P(o_t \mid s_t)$$

Sequence Likelihood

- Joint probability

P( )



1. Markov assumption:

$$P(s_{t+1} \mid s_1, \ldots, s_t) \approx P(s_{t+1} \mid s_t)$$

2. Output independence:

$$P(o_t \mid s_1, \ldots, s_t) \approx P(o_t \mid s_t)$$

Learning the parameters of HMM

Training set:

1 Pierre/NNP Vinken/NNP ,/, 61/CD years/NNS old/JJ ,/
join/VB the/DT board/NN as/IN a/DT nonexecutive/JJ di
Nov./NNP 29/CD ./.
2 Mr./NNP Vinken/NNP is/VBZ chairman/NN of/IN Elsev
N.V./NNP ,/, the/DT Dutch/NNP publishing/VBG group/
3 Rudolph/NNP Agnew/NNP ,/, 55/CD years/NNS old/JJ
chairman/NN of/IN Consolidated/NNP Gold/NNP Fields/N
,/, was/VBD named/VBN a/DT nonexecutive/JJ director/
this/DT British/JJ industrial/JJ conglomerate/NN ./.
...
38,219 It/PRP is/VBZ also/RB pulling/VBG 20/CD peopl
of/IN Puerto/NNP Rico/NNP ,/, who/WP were/VBD help
Huricane/NNP Hugo/NNP victims/NNS ,/, and/CC sendin
them/PRP to/TO San/NNP Francisco/NNP instead/RB ./

- Maximum likelihood estimate:

$$P(s_i \mid s_j) = \frac{Count(s_j, s_i)}{Count(s_j)}$$

$$P(o \mid s) = \frac{Count(s, o)}{Count(s)}$$

Learning the parameters of HMM

1. the/DT cat/NN sat/VBD on/IN the/DT mat/NN

2. Princeton/NNP is/VBZ in/IN New/NNP Jersey/NNP

3. the/DT old/NN man/VB the/DT boats/NNS

$$P(NN \mid DT) = \frac{3}{4}$$

$$P(cat \mid NN) = \frac{1}{3}$$

- Maximum likelihood estimate:

$$P(s_i \mid s_j) = \frac{Count(s_j, s_i)}{Count(s_j)}$$

$$P(o \mid s) = \frac{Count(s, o)}{Count(s)}$$

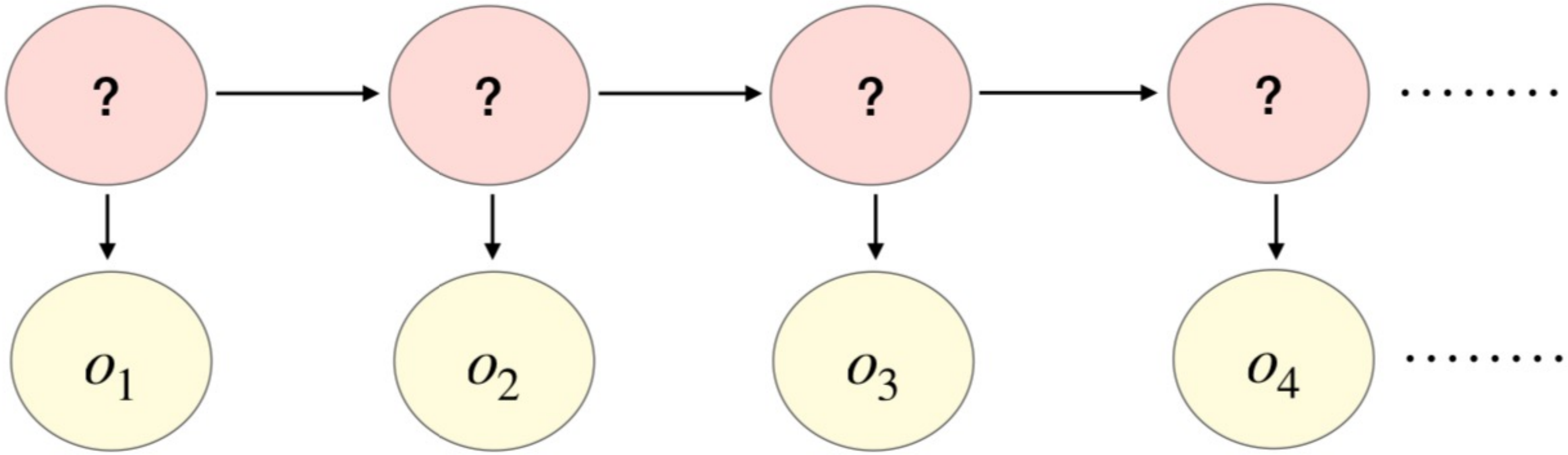
	N	M	V
mark	4/9	0	0
will	1/9	3/4	0
mary	2/9	0	0
park	2/9	0	1/4
can	0	1/4	0
see	0	0	1/2
support	0	0	1/4

	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	2/3	0	0	1/3

<S> N M V <E>

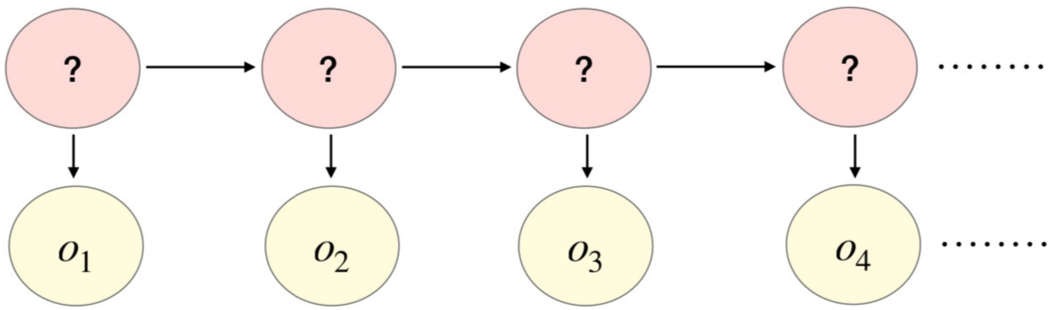
P(Mark/N will/M park/V)

Decoding with HMM

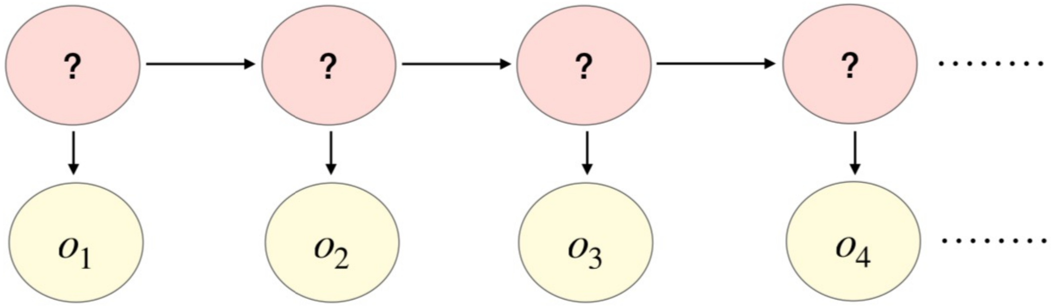


Given the observation (a sequence of words), find the most likely sequence of states (s sequence of pos tags)

Decoding with HMM



Decoding with HMM



- Viterbi algorithm
  - Dynamic programming

Summary

- POS tagging:
  - Input: a sequence of word tokens **w**
  - Output: a sequence of tags **t**, one per word (t$\in \mathcal{T}$)

- Why do POS tagging

- Computing sequence likelihood POS tagging