Sequence Models I

This Lecture

- Sequence modeling
- HMMs for POS tagging
- HMM parameter estimation
- Viterbi

POS Tagging

- What tags are out there?

Ghana 's ambassador should have set up the big meeting in DC yesterday .

POS Tagging

NNP VBZ NN NNS CD NN
Fed raises interest rates 0.5 percent

PRP VBP CD NN IN DT NNP NNS
I have 0.5% interest in the Fed raises

NNP - Proper Noun, Singular
NNS - Proper Noun, Plural
NN - Common Noun
PRP - Pronoun, Personal
CD- Numerical
VBZ- Verb Present Tens, 3rd Person Singular
VBP - Verb Present Tens, not 3rd Person Singular
IN- Preposition/Conjunction
DT- Determiner

- How to choose the correct tag from the tagset?

POS Tagging

CC	conjunction, coordinating	and both but either or
CD	numeral, cardinal	mid-1890 nine-thirty 0.5 one
DT	determiner	a all an every no that the
EX	existential there	there
FW	foreign word	gemeinschaft hund ich jeux
IN	preposition or conjunction, subordinating	among whether out on by if
JJ	adjective or numeral, ordinal	third ill-mannered regrettable
JJR	adjective, comparative	braver cheaper taller
JJS	adjective, superlative	bravest cheapest tallest
MD	modal auxiliary	can may might will would
NN	noun, common, singular or mass	cabbage thermostat investment subhumanity
NNP	noun, proper, singular	Motown Cougar Yvette Liverpool
NNPS	noun, proper, plural	Americans Materials States
NNS	noun, common, plural	undergraduates bric-a-brac averages
POS	genitive marker	' 's
PRP	pronoun, personal	hers himself it we them
PRP$	pronoun, possessive	her his mine my our ours their thy your
RB	adverb	occasionally maddeningly adventurously
RBR	adverb, comparative	further gloomier heavier less-perfectly
RBS	adverb, superlative	best biggest nearest worst
RP	particle	aboard away back by on open through
TO	"to" as preposition or infinitive marker	to
UH	interjection	huh howdy uh whammo shucks heck
VB	verb, base form	ask bring fire see take
VBD	verb, past tense	pleaded swiped registered saw
VBG	verb, present participle or gerund	stirring focusing approaching erasing
VBN	verb, past participle	dilapidated imitated reunifed unsettled
VBP	verb, present tense, not 3rd person singular	twist appear comprise mold postpone
VBZ	verb, present tense, 3rd person singular	bases reconstructs marks uses
WDT	WH-determiner	that what whatever which whichever
WP	WH-pronoun	that what whatever which who whom
WP$	WH-pronoun, possessive	whose
WRB	Wh-adverb	however whenever where why

What is this good for?

What is this good for?

- Text-to-speech: *record*, *lead*

What is this good for?

- Text-to-speech: *record*, *lead*
- Preprocessing step for syntactic parsers

What is this good for?

- Text-to-speech: *record*, *lead*
- Preprocessing step for syntactic parsers
- Domain-independent disambiguation for other tasks

What is this good for?

- Text-to-speech: *record*, *lead*
- Preprocessing step for syntactic parsers
- Domain-independent disambiguation for other tasks
- (Very) shallow information extraction

Sequence Models

Sequence Models

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$

- POS tagging: ***x*** is a sequence of words, ***y*** is a sequence of tags

- Today: generative models P(***x, y***); discriminative models next time

Hidden Markov Models

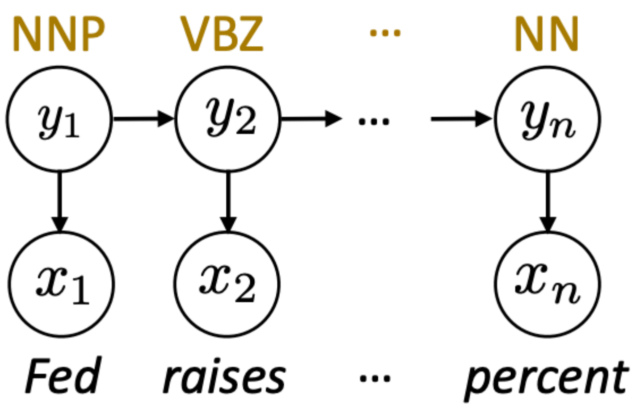
- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$
- Model the sequence of *y* as a Markov process (dynamics model)
- Markov property: future is conditionally independent of the past given the present

$$y_1 \rightarrow y_2 \rightarrow y_3 \qquad P(y_3|y_1, y_2) = P(y_3|y_2)$$

- Lots of mathematical theory about how Markov chains behave
- If *y* are tags, this roughly corresponds to assuming that the next tag only depends on the current tag, not anything before

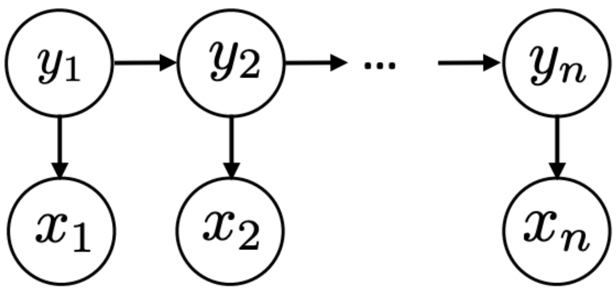
Hidden Markov Models

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$



Hidden Markov Models

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$



$$P(\mathbf{y},\mathbf{x}) = \underbrace{P(y_1)}_{\text{Initial distribution}} \underbrace{\prod_{i=2}^{n} P(y_i|y_{i-1})}_{\text{Transition probabilities}} \underbrace{\prod_{i=1}^{n} P(x_i|y_i)}_{\text{Emission probabilities}}$$

- Observation (*x*) depends only on current state (*y*)
- Multinomials: tag x tag transitions, tag x word emissions
- P(*x*|*y*) is a distribution over all words in the vocabulary — not a distribution over features (but could be!)

Transitions in POS Tagging

- Dynamics model $P(y_1)\prod_{i=2}^{n} P(y_i|y_{i-1})$

VBD		VB				
VBN	VBZ	VBP	VBZ			
NNP	NNS	NN	NNS	CD	NN	.
Fed	*raises*	*interest*	*rates*	*0.5*	*percent*	*.*

NNP - proper noun, singular
VBZ - verb, 3rd ps. sing. present
NN - noun, singular or mass

- $P(y_1 = \mathrm{NNP})$ likely because start of sentence
- $P(y_2 = \mathrm{VBZ}|y_1 = \mathrm{NNP})$ likely because verb often follows noun
- $P(y_3 = \mathrm{NN}|y_2 = \mathrm{VBZ})$ direct object follows verb, other verb rarely follows past tense verb (main verbs can follow modals though!)

Estimating Transitions

NNP VBZ NN NNS CD NN .
Fed raises interest rates 0.5 percent .

- Similar to Naive Bayes estimation: maximum likelihood solution = normalized counts (with smoothing) read off supervised data
- P(tag | NN) = (0.5 ., 0.5 NNS)
- How to smooth?
- One method: smooth with unigram distribution over tags

$$P(\mathrm{tag}|\mathrm{tag}_{-1}) = (1-\lambda)\hat{P}(\mathrm{tag}|\mathrm{tag}_{-1}) + \lambda\hat{P}(\mathrm{tag})$$

$\hat{P}$ = empirical distribution (read off from data)

Emissions in POS Tagging

NNP VBZ NN NNS CD NN .
Fed raises interest rates 0.5 percent .

- Emissions P(*x* | *y*) capture the distribution of words occurring with a given tag
- P(word | NN) = (0.05 *person*, 0.04 *official*, 0.03 *interest*, 0.03 *percent* …)
- When you compute the posterior for a given word's tags, the distribution favors tags that are more likely to generate that word
- How should we smooth this?

Emissions in POS Tagging

NNP VBZ NN NNS CD NN
Fed raises interest rates 0.5 percent

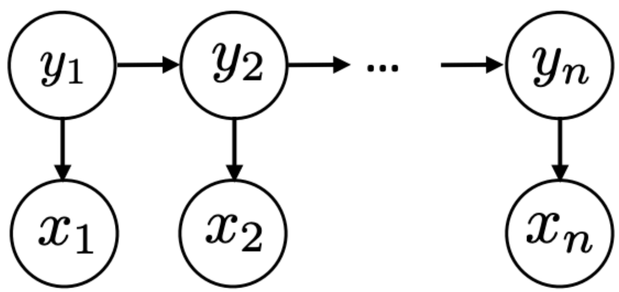
- P(word | NN) = (0.5 *interest*, 0.5 *percent*) — hard to smooth!
- Can interpolate with distribution looking at word shape P(word shape | tag) (e.g., P(capitalized word of len >= 8 | tag))
- Alternative: use Bayes' rule

$$P(\mathrm{word}|\mathrm{tag}) = \frac{P(\mathrm{tag}|\mathrm{word})P(\mathrm{word})}{P(\mathrm{tag})}$$

  - Fancy techniques from language modeling, e.g. look at type fertility — P(tag|word) is flatter for some kinds of words than for others)
- P(word|tag) can be a log-linear model — we'll see this in a few lectures

Inference in HMMs

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$

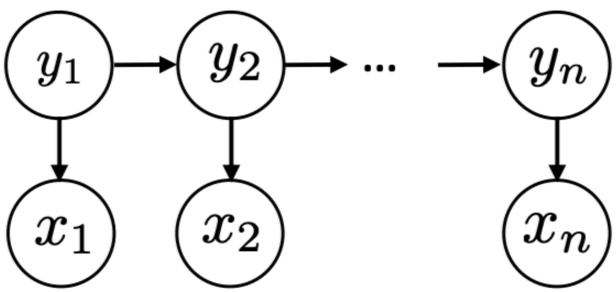


$$P(\mathbf{y}, \mathbf{x}) = P(y_1) \prod_{i=2}^{n} P(y_i | y_{i-1}) \prod_{i=1}^{n} P(x_i | y_i)$$

- Inference problem: $\mathrm{argmax}_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) = \mathrm{argmax}_{\mathbf{y}} \frac{P(\mathbf{y}, \mathbf{x})}{P(\mathbf{x})}$

Inference in HMMs

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$

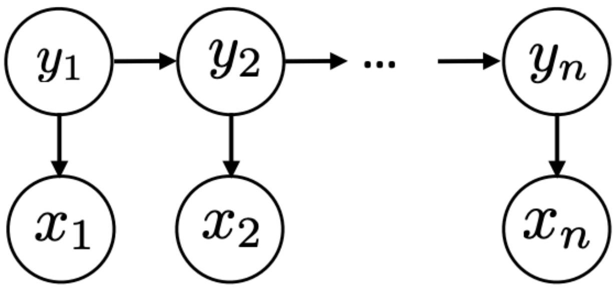


$$P(\mathbf{y}, \mathbf{x}) = P(y_1) \prod_{i=2}^{n} P(y_i | y_{i-1}) \prod_{i=1}^{n} P(x_i | y_i)$$

- Inference problem: $\text{argmax}_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) = \text{argmax}_{\mathbf{y}} \frac{P(\mathbf{y}, \mathbf{x})}{\cancel{P(\mathbf{x})}}$
- Exponentially many possible ***y*** here!

Inference in HMMs

- Input $\mathbf{x} = (x_1, ..., x_n)$ Output $\mathbf{y} = (y_1, ..., y_n)$



$$P(\mathbf{y}, \mathbf{x}) = P(y_1) \prod_{i=2}^{n} \boxed{P(y_i | y_{i-1})} \prod_{i=1}^{n} P(x_i | y_i)$$

- Inference problem: $\mathrm{argmax}_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) = \mathrm{argmax}_{\mathbf{y}} \frac{P(\mathbf{y}, \mathbf{x})}{\cancel{P(\mathbf{x})}}$
- Exponentially many possible ***y*** here!
- Solution: dynamic programming (possible because of Markov structure!)

Example

Sentence: “Learning changes people”
Tagset: NN, VB

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$

$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$

$P(q_f|VB) = 2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

VB			
NN			
Input	Learning	changes	People

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

VB			
NN	$2*10^{-4}$ (q_0)		
Input	Learning	changes	People

$P(Learning|VB)= 3*10^{-3}$
$P(Learning|NN)= 1*10^{-3}$
$P(People|NN)= 5*10^{-2}$
$P(People|VB)= 2*10^{-4}$
$P(changes|NN)= 3*10^{-3}$
$P(changes|VB)= 4*10^{-2}$

$P(NN|VB)= 4*10^{-1}$
$P(NN|NN)= 1*10^{-1}$
$P(VB|VB)= 1*10^{-1}$
$P(VB|NN)= 3*10^{-1}$

$P(VB|q_0)= 3*10^{-1}$
$P(NN|q_0)= 2*10^{-1}$

$P(q_f|NN)=1*10^{-1}$
$P(q_f|VB)= 2*10^{-2}$

$P(Learning|NN) = P(Learning|NN) * P(NN|q_0)= 2*10^{-4}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)		
NN	$2*10^{-4}$ (q_0)		
Input	Learning	changes	People

$P(Learning|VB)= 3*10^{-3}$
$P(Learning|NN)= 1*10^{-3}$
$P(People|NN)= 5*10^{-2}$
$P(People|VB)= 2*10^{-4}$
$P(changes|NN)= 3*10^{-3}$
$P(changes|VB)= 4*10^{-2}$

$P(NN|VB)= 4*10^{-1}$
$P(NN|NN)= 1*10^{-1}$
$P(VB|VB)= 1*10^{-1}$
$P(VB|NN)= 3*10^{-1}$

$P(VB|q_0)= 3*10^{-1}$
$P(NN|q_0)= 2*10^{-1}$

$P(q_f|NN)=1*10^{-1}$
$P(q_f|VB)= 2*10^{-2}$

$P(Learning|NN) = P(Learning|NN) * P(NN|q_0)= 2*10^{-4}$

$P(Learning|VB) = P(Learning|VB) * P(VB|q_0)= 9*10^{-4}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)		
NN	$2*10^{-4}$ (q_0)		
Input	Learning	changes	People

$P(Learning|VB) = 3*10^{-3}$
$P(Learning|NN) = 1*10^{-3}$
$P(People|NN) = 5*10^{-2}$
$P(People|VB) = 2*10^{-4}$
$P(changes|NN) = 3*10^{-3}$
$P(changes|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

$P(changes|NN) = P(changes|NN) * P(NN|NN) * 2*10^{-4} = 6*10^{-8}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)		
NN	$2*10^{-4}$ (q_0)		
Input	Learning	changes	People

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

$P(\text{changes}|NN) = P(\text{changes}|NN) * P(NN|NN) * 2*10^{-4} = 6*10^{-8}$

$P(\text{changes}|NN) = P(\text{changes}|NN) * P(NN|VB) * 9*10^{-4} = 1.08*10^{-6}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)		
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	
Input	Learning	changes	People

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

$P(\text{changes}|NN) = P(\text{changes}|NN) * P(NN|NN) * 2*10^{-4} = 6*10^{-8}$

$P(\text{changes}|NN) = P(\text{changes}|NN) * P(NN|VB) * 9*10^{-4} = 1.08*10^{-6}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)		
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	
Input	Learning	changes	People

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

$P(\text{changes}|VB) = P(\text{changes}|VB) * P(VB|NN) * 2*10^{-4} = 2.4*10^{-6}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(changes|VB) = P(changes|VB) * P(VB|NN) * $2*10^{-4}$ = $2.4*10^{-6}$

P(changes|VB) = P(changes|VB) * P(VB|VB) * $9*10^{-4}$ = $3.6*10^{-6}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0) →	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)= $1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(changes|VB) = P(changes|VB) * P(VB|NN) * $2*10^{-4} = 2.4*10^{-6}$

P(changes|VB) = P(changes|VB) * P(VB|VB) * $9*10^{-4} = 3.6*10^{-6}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|NN) = P(People|NN) * P(NN|NN) * $1.08*10^{-6}$ = $5.4*10^{-9}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|NN) = P(People|NN) * P(NN|NN) * $1.08*10^{-6}$ = $5.4*10^{-9}$

P(People|NN) = P(People|NN) * P(NN|VB) * $3.6*10^{-6}$ = $7.2*10^{-8}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

$P(Learning|VB) = 3*10^{-3}$
$P(Learning|NN) = 1*10^{-3}$
$P(People|NN) = 5*10^{-2}$
$P(People|VB) = 2*10^{-4}$
$P(changes|NN) = 3*10^{-3}$
$P(changes|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

$P(People|NN) = P(People|NN) * P(NN|NN) * 1.08*10^{-6} = 5.4*10^{-9}$

$P(People|NN) = P(People|NN) * P(NN|VB) * 3.6*10^{-6} = 7.2*10^{-8}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	$7.2*10^{-11}$ (VB)
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|VB) = P(People|VB) * P(VB|NN) * $1.08*10^{-6}$ = $6.48*10^{-11}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|VB) = P(People|VB) * P(VB|NN) * $1.08*10^{-6}$ = $6.48*10^{-11}$

P(People|VB) = P(People|VB) * P(VB|VB) * $3.6*10^{-6}$ = $7.2*10^{-11}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	$7.2*10^{-11}$ (VB)
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|VB) = P(People|VB) * P(VB|NN) * $1.08*10^{-6}$ = $6.48*10^{-11}$

P(People|VB) = P(People|VB) * P(VB|VB) * $3.6*10^{-6}$ = $7.2*10^{-11}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	$7.2*10^{-11}$ (VB)
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

$P(\text{Learning}|VB)= 3*10^{-3}$
$P(\text{Learning}|NN)= 1*10^{-3}$
$P(\text{People}|NN)= 5*10^{-2}$
$P(\text{People}|VB)= 2*10^{-4}$
$P(\text{changes}|NN)= 3*10^{-3}$
$P(\text{changes}|VB)= 4*10^{-2}$

$P(NN|VB)= 4*10^{-1}$
$P(NN|NN)= 1*10^{-1}$
$P(VB|VB)= 1*10^{-1}$
$P(VB|NN)= 3*10^{-1}$

$P(VB|q_0)= 3*10^{-1}$
$P(NN|q_0)= 2*10^{-1}$

$P(q_f|NN)=1*10^{-1}$
$P(q_f|VB)= 2*10^{-2}$

$P(\text{People}|VB) = P(\text{People}|VB) * P(VB|NN) * 1.08*10^{-6} = 6.48*10^{-11}$

$P(\text{People}|VB) = P(\text{People}|VB) * P(VB|VB) * 3.6*10^{-6} = 7.2*10^{-11}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

q_F			
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-6}$ (VB)	$7.2*10^{-11}$ (VB)
NN	$2*10^{-4}$ (q_0)	$1.08*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)
Input	Learning	changes	People

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

P(People|VB) = P(People|VB) * P(VB|NN) * $1.08*10^{-6}$ = $6.48*10^{-11}$

P(People|VB) = P(People|VB) * P(VB|VB) * $3.6*10^{-6}$ = $7.2*10^{-11}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

					q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

$P(q_F|VB) = P(q_f|VB) * 7.2*10^{-11} = 1.44*10^{-12}$

$P(Learning|VB) = 3*10^{-3}$
$P(Learning|NN) = 1*10^{-3}$
$P(People|NN) = 5*10^{-2}$
$P(People|VB) = 2*10^{-4}$
$P(changes|NN) = 3*10^{-3}$
$P(changes|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

$P(q_F|VB) = P(q_f|VB) * 7.2*10^{-11} = 1.44*10^{-12}$

$P(q_F|NN) = P(q_f|NN) * 7.2*10^{-8} = 7.2*10^{-9}$

$P(\text{Learning}|VB) = 3*10^{-3}$
$P(\text{Learning}|NN) = 1*10^{-3}$
$P(\text{People}|NN) = 5*10^{-2}$
$P(\text{People}|VB) = 2*10^{-4}$
$P(\text{changes}|NN) = 3*10^{-3}$
$P(\text{changes}|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

Example

Sentence: “Learning changes people”
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

$P(q_F|VB) = P(q_f|VB) * 7.2*10^{-11} = 1.44*10^{-12}$

$P(q_F|NN) = P(q_f|NN) * 7.2*10^{-8} = 7.2*10^{-9}$

$P(Learning|VB) = 3*10^{-3}$
$P(Learning|NN) = 1*10^{-3}$
$P(People|NN) = 5*10^{-2}$
$P(People|VB) = 2*10^{-4}$
$P(changes|NN) = 3*10^{-3}$
$P(changes|VB) = 4*10^{-2}$

$P(NN|VB) = 4*10^{-1}$
$P(NN|NN) = 1*10^{-1}$
$P(VB|VB) = 1*10^{-1}$
$P(VB|NN) = 3*10^{-1}$

$P(VB|q_0) = 3*10^{-1}$
$P(NN|q_0) = 2*10^{-1}$

$P(q_f|NN) = 1*10^{-1}$
$P(q_f|VB) = 2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

$P(\text{Learning}|VB)= 3*10^{-3}$
$P(\text{Learning}|NN)= 1*10^{-3}$
$P(\text{People}|NN)= 5*10^{-2}$
$P(\text{People}|VB)= 2*10^{-4}$
$P(\text{changes}|NN)= 3*10^{-3}$
$P(\text{changes}|VB)= 4*10^{-2}$

$P(NN|VB)= 4*10^{-1}$
$P(NN|NN)= 1*10^{-1}$
$P(VB|VB)= 1*10^{-1}$
$P(VB|NN)= 3*10^{-1}$

$P(VB|q_0)= 3*10^{-1}$
$P(NN|q_0)= 2*10^{-1}$

$P(q_f|NN)=1*10^{-1}$
$P(q_f|VB)= 2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

Example

Sentence: "Learning changes people"
Tagset: NN, VB

				q_F
VB	$9*10^{-4}$ (q_0)	$3.6*10^{-7}$ (VB)	$7.2*10^{-11}$ (VB)	$1.44*10^{-12}$ (VB)
NN	$2*10^{-4}$ (q_0)	$10.8*10^{-6}$ (VB)	$7.2*10^{-8}$ (VB)	$7.2*10^{-9}$ (NN)
Input	Learning	changes	People	
Output	VB	VB	NN	

P(Learning|VB)= $3*10^{-3}$
P(Learning|NN)= $1*10^{-3}$
P(People|NN)= $5*10^{-2}$
P(People|VB)= $2*10^{-4}$
P(changes|NN)= $3*10^{-3}$
P(changes|VB)= $4*10^{-2}$

P(NN|VB)= $4*10^{-1}$
P(NN|NN)= $1*10^{-1}$
P(VB|VB)= $1*10^{-1}$
P(VB|NN)= $3*10^{-1}$

P(VB|q_0)= $3*10^{-1}$
P(NN|q_0)= $2*10^{-1}$

P(q_f|NN)=$1*10^{-1}$
P(q_f|VB)= $2*10^{-2}$

Viterbi Algorithm

1. Initial: For each state s, calculate

$$\text{score}_1(s) = P(s)P(x_1|s) = \pi_s B_{x_1,s}$$

2. Recurrence: For i = 2 to n, for every state s, calculate

$$\text{score}_i(s) = \max_{y_{i-1}} P(s|y_{i-1})P(x_i|s)\text{score}_{i-1}(y_{i-1})$$
$$= \max_{y_{i-1}} A_{y_{i-1},s} B_{s,x_i} \text{score}_{i-1}(y_{i-1})$$

3. Final state: calculate

$$\max_{\mathbf{y}} P(\mathbf{y}, \mathbf{x}|\pi, A, B) = \max_{s} \text{score}_n(s)$$

π: Initial probabilities
A: Transitions
B: Emissions

This only calculates the max. To get final answer (*argmax*),

- keep track of which state corresponds to the max at each step
- build the answer using these back pointers