Part of Speech Tagging

Natural Language Processing: Jordan Boyd-Graber

University of Maryland

Viterbi

Adapted from material by Jimmy Lin and Jason Eisner

Viterbi Algorithm

- Given an unobserved sequence of length L, $\{x_1, \ldots, x_L\}$, we want to find a sequence $\{z_1 \ldots z_L\}$ with the highest probability.

Viterbi Algorithm

- Given an unobserved sequence of length L, $\{x_1, \ldots, x_L\}$, we want to find a sequence $\{z_1 \ldots z_L\}$ with the highest probability.
- It's impossible to compute K^L possibilities.
- So, we use dynamic programming to compute most likely tags for each token subsequence from 0 to t that ends in state k.
- Memoization: fill a table of solutions of sub-problems
- Solve larger problems by composing sub-solutions
- Base case:

$$\delta_1(k) = \pi_k \beta_{k,x_i} \tag{1}$$

- Recursion:

$$\delta_n(k) = \max_j \left(\delta_{n-1}(j)\theta_{j,k}\right)\beta_{k,x_n} \tag{2}$$

Viterbi Algorithm

- Given an unobserved sequence of length L, $\{x_1, \dots, x_L\}$, we want to find a sequence $\{z_1 \dots z_L\}$ with the highest probability.
- It's impossible to compute K^L possibilities.
- So, we use dynamic programming to compute most likely tags for each token subsequence from 0 to t that ends in state k.
- Memoization: fill a table of solutions of sub-problems
- Solve larger problems by composing sub-solutions
- Base case:

$$\delta_1(k) = \pi_k \beta_{k,x_i} \tag{1}$$

- Recursion:

$$\delta_n(k) = \max_{j} \left(\delta_{n-1}(j) \theta_{j,k} \right) \beta_{k,x_n} \tag{2}$$

Viterbi Algorithm

- Given an unobserved sequence of length L, $\{x_1, \ldots, x_L\}$, we want to find a sequence $\{z_1 \ldots z_L\}$ with the highest probability.
- It's impossible to compute K^L possibilities.
- So, we use dynamic programming to compute most likely tags for each token subsequence from 0 to t that ends in state k.
- Memoization: fill a table of solutions of sub-problems
- Solve larger problems by composing sub-solutions
- Base case:
$$\delta_1(k) = \pi_k \beta_{k,x_i} \tag{1}$$
- Recursion:
$$\delta_n(k) = \max_j \left(\delta_{n-1}(j) \theta_{j,k}\right) \beta_{k,x_n} \tag{2}$$

Viterbi Algorithm

- Given an unobserved sequence of length L, $\{x_1, \ldots, x_L\}$, we want to find a sequence $\{z_1 \ldots z_L\}$ with the highest probability.
- It's impossible to compute K^L possibilities.
- So, we use dynamic programming to compute most likely tags for each token subsequence from 0 to t that ends in state k.
- Memoization: fill a table of solutions of sub-problems
- Solve larger problems by composing sub-solutions
- Base case:

$$\delta_1(k) = \pi_k \beta_{k,x_i} \tag{1}$$

- Recursion:

$$\delta_n(k) = \max_j \left(\delta_{n-1}(j)\theta_{j,k}\right)\beta_{k,x_n} \tag{2}$$

- The complexity of this is now K^2L.
- In class: example that shows why you need all $O(KL)$ table cells (garden pathing)
- But just computing the max isn't enough. We also have to remember where we came from. (Breadcrumbs from best previous state.)

$$\Psi_n = \text{argmax}_j \delta_{n-1}(j)\theta_{j,k} \tag{3}$$

- The complexity of this is now $K^2 L$.
- In class: example that shows why you need all $O(KL)$ table cells (garden pathing)
- But just computing the max isn't enough. We also have to remember where we came from. (Breadcrumbs from best previous state.)

$$\Psi_n = \text{argmax}_j \delta_{n-1}(j) \theta_{j,k} \tag{3}$$

- Let's do that for the sentence "come and get it"

POS	π_k	β_{k,x_1}	$\log \delta_1(k)$
MOD	0.234	0.024	-5.18
DET	0.234	0.032	-4.89
CONJ	0.234	0.024	-5.18
N	0.021	0.016	-7.99
PREP	0.021	0.024	-7.59
PRO	0.021	0.016	-7.99
V	0.234	0.121	-3.56

come and get it

Why logarithms?

1. More interpretable than a float with lots of zeros.
2. Underflow is less of an issue
3. Addition is cheaper than multiplication

$$log(ab) = log(a) + log(b) \tag{4}$$

POS	$\log \delta_1(j)$		$\log \delta_2(\text{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		
N	-7.99		
PREP	-7.59		
PRO	-7.99		
V	-3.56		

come **and** get it

POS	$\log \delta_1(j)$		$\log \delta_2(\mathrm{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		???
N	-7.99		
PREP	-7.59		
PRO	-7.99		
V	-3.56		

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		???
N	-7.99		
PREP	-7.59		
PRO	-7.99		
V	-3.56		

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		???
N	-7.99		
PREP	-7.59		
PRO	-7.99		
V	-3.56		

come **and** get it

$$\log\left(\delta_0(\text{V})\theta_{\text{V, CONJ}}\right) = \log \delta_0(k) + \log \theta_{\text{V, CONJ}} = -3.56 + -1.65$$

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\mathrm{CONJ}}$	$\log \delta_2(\mathrm{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		???
N	-7.99		
PREP	-7.59		
PRO	-7.99		
V	-3.56	-5.21	

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18		
DET	-4.89		
CONJ	-5.18		???
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18	-8.48	
DET	-4.89	-7.72	
CONJ	-5.18	-8.47	???
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18	-8.48	
DET	-4.89	-7.72	
CONJ	-5.18	-8.47	???
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\mathrm{CONJ}}$	$\log \delta_2(\mathrm{CONJ})$
MOD	-5.18	-8.48	
DET	-4.89	-7.72	
CONJ	-5.18	-8.47	
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

$\log \delta_1(k) = -5.21 - \log \beta_{\mathrm{CONJ}}$, and $=$

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\mathrm{CONJ}}$	$\log \delta_2(\mathrm{CONJ})$
MOD	-5.18	-8.48	
DET	-4.89	-7.72	
CONJ	-5.18	-8.47	
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

$$\log \delta_1(k) = -5.21 - \log \beta_{\mathrm{CONJ,\ and}} = -5.21 - 0.64$$

POS	$\log \delta_1(j)$	$\log \delta_1(j)\theta_{j,\text{CONJ}}$	$\log \delta_2(\text{CONJ})$
MOD	-5.18	-8.48	
DET	-4.89	-7.72	
CONJ	-5.18	-8.47	-6.02
N	-7.99	≤ -7.99	
PREP	-7.59	≤ -7.59	
PRO	-7.99	≤ -7.99	
V	-3.56	-5.21	

come **and** get it

POS	$\delta_1(k)$	$\delta_2(k)$	b_2	$\delta_3(k)$	b_3	$\delta_4(k)$	b_4
MOD	-5.18						
DET	-4.89						
CONJ	-5.18	-6.02	V				
N	-7.99						
PREP	-7.59						
PRO	-7.99						
V	-3.56						
WORD	come	and		get		it	

POS	$\delta_1(k)$	$\delta_2(k)$	b_2	$\delta_3(k)$	b_3	$\delta_4(k)$	b_4
MOD	-5.18	-0.00	X				
DET	-4.89	-0.00	X				
CONJ	-5.18	-6.02	V				
N	-7.99	-0.00	X				
PREP	-7.59	-0.00	X				
PRO	-7.99	-0.00	X				
V	-3.56	-0.00	X				
WORD	come	and		get		it	

POS	$\delta_1(k)$	$\delta_2(k)$	b_2	$\delta_3(k)$	b_3	$\delta_4(k)$	b_4
MOD	-5.18	-0.00	X	-0.00	X		
DET	-4.89	-0.00	X	-0.00	X		
CONJ	-5.18	-6.02	V	-0.00	X		
N	-7.99	-0.00	X	-0.00	X		
PREP	-7.59	-0.00	X	-0.00	X		
PRO	-7.99	-0.00	X	-0.00	X		
V	-3.56	-0.00	X	-9.03	CONJ		
WORD	come	and		get		it	

POS	$\delta_1(k)$	$\delta_2(k)$	b_2	$\delta_3(k)$	b_3	$\delta_4(k)$	b_4
MOD	-5.18	-0.00	X	-0.00	X	-0.00	X
DET	-4.89	-0.00	X	-0.00	X	-0.00	X
CONJ	-5.18	-6.02	V	-0.00	X	-0.00	X
N	-7.99	-0.00	X	-0.00	X	-0.00	X
PREP	-7.59	-0.00	X	-0.00	X	-0.00	X
PRO	-7.99	-0.00	X	-0.00	X	-14.6	V
V	-3.56	-0.00	X	-9.03	CONJ	-0.00	X
WORD	come	and		get		it	

What if you don't have training data?

- You can still learn a hmm
- Using a general technique called expectation maximization

What if you don't have training data?

- You can still learn a hmm
- Using a general technique called expectation maximization
  - Take a guess at the parameters
  - Figure out latent variables
  - Find the parameters that best explain the latent variables
  - Repeat

em for hmm

Model Parameters

We need to start with model parameters

em for hmm

Model Parameters

π, β, θ

We can initialize these any way we want

em for hmm

Model Parameters

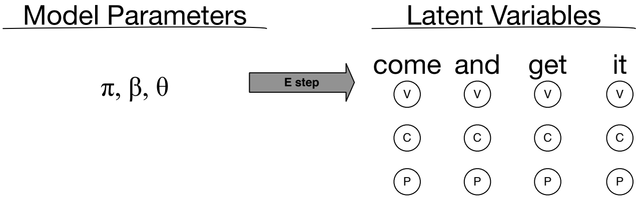
$\pi, \beta, \theta \xrightarrow{\text{E step}}$

em for hmm

Model Parameters		Latent Variables
π, β, θ	$\xrightarrow{\text{E step}}$	come and get it

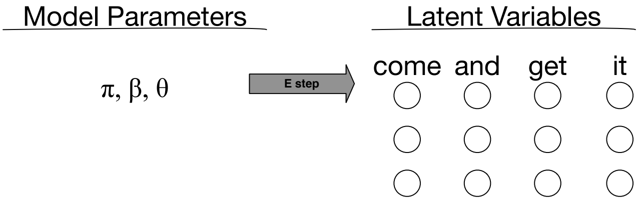
We compute the E-step based on our data

em for hmm



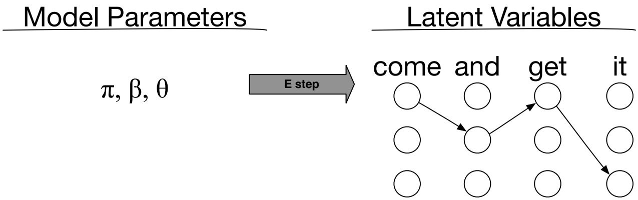
Each word in our dataset could take any part of speech

em for hmm



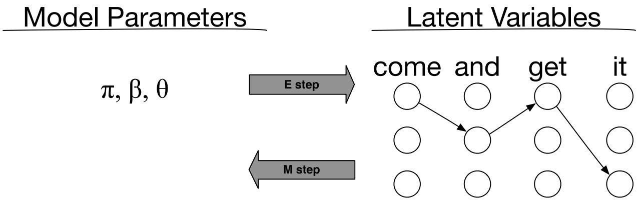
But we don’t know which state was used for each word

em for hmm



Determine the probability of being in each latent state using Forward / Backward

em for hmm

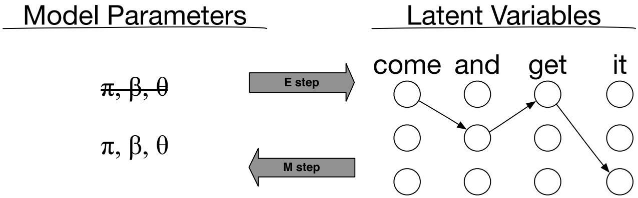


Calculate new parameters:

$$\theta_i = \frac{n_i + \alpha_i}{\sum_k \mathbb{E}_p[n_k] + \alpha_k} \tag{5}$$

Where the expected counts are from the lattice

em for hmm



Replace old parameters (and start over)

Hard vs. Full EM

Hard EM

Train only on the most likely sentence (Viterbi)

- Faster: E-step is faster
- Faster: Fewer iterations

Full EM

Compute probability of all possible sequences

- More accurate: Doesn't get stuck in local optima as easily

Warning about next homework(s)

- Kaggle competition
- Thus, late days not very useful
- Following homework is not computational

Garden Pathing

What is the probability of the sequence “a/`Det` blue/`Adj` boat/`N`”?

Garden Pathing

What is the probability of the sequence "a/`Det` blue/`Adj` boat/`N`"?

$$\pi_d \beta_{d,the} \theta_{d,a} \beta_{a,blue} \theta_{a,n} \beta_{n,boat} = \tag{5}$$

$$0.3 * 0.6 * 0.4 * 0.3 * 0.5 * 0.1 = 0.00108 \tag{6}$$

Garden Pathing

What is the probability of the sequence "a/`Det` blue/`Adj` boat/`N`"?

$$\pi_d \beta_{d,the} \theta_{d,a} \beta_{a,blue} \theta_{a,n} \beta_{n,boat} = \tag{5}$$

$$0.3 * 0.6 * 0.4 * 0.3 * 0.5 * 0.1 = 0.00108 \tag{6}$$

Example . . .

Base case

Example ...

Base case

1. $\delta_1(a) = -4.6$

Example ...

Base case

1. $\delta_1(a) = -4.6$
2. $\delta_1(v) = -5.7$

Example . . .

Base case

1. $\delta_1(a) = -4.6$
2. $\delta_1(v) = -5.7$
3. $\delta_1(d) = -1.7$

Example ...

Base case

1. $\delta_1(a) = -4.6$
2. $\delta_1(v) = -5.7$
3. $\delta_1(d) = -1.7$
4. $\delta_1(n) = -4.6$

Example ...

Second position

1. $\delta_2(a) = \max\left(\underbrace{-5.8}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-2.6}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -1.2 = -2.6 + -1.2 = -3.8$

Example ...

Second position

1. $\delta_2(a) = \max\left(\underbrace{-5.8}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-2.6}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -1.2 = -2.6 + -1.2 = -3.8$
2. $\delta_2(v) = \max\left(\underbrace{-6.9}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-4.7}_{\textbf{d}}, \underbrace{-4.8}_{\text{n}}\right) + -2.3 = -4.7 + -2.3 = -7.0$

Example ...

Second position

1. $\delta_2(a) = \max\left(\underbrace{-5.8}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-2.6}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -1.2 = -2.6 + -1.2 = -3.8$
2. $\delta_2(v) = \max\left(\underbrace{-6.9}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-4.7}_{\textbf{d}}, \underbrace{-4.8}_{\text{n}}\right) + -2.3 = -4.7 + -2.3 = -7.0$
3. $\delta_2(d) = \max\left(\underbrace{-6.9}_{\text{a}}, \underbrace{-6.9}_{\text{v}}, \underbrace{-4.0}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -3.7 = -4.0 + -3.7 = -7.7$

Example ...

Second position

1. $\delta_2(a) = \max\left(\underbrace{-5.8}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-2.6}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -1.2 = -2.6 + -1.2 = -3.8$
2. $\delta_2(v) = \max\left(\underbrace{-6.9}_{\text{a}}, \underbrace{-7.3}_{\text{v}}, \underbrace{-4.7}_{\textbf{d}}, \underbrace{-4.8}_{\text{n}}\right) + -2.3 = -4.7 + -2.3 = -7.0$
3. $\delta_2(d) = \max\left(\underbrace{-6.9}_{\text{a}}, \underbrace{-6.9}_{\text{v}}, \underbrace{-4.0}_{\textbf{d}}, \underbrace{-7.6}_{\text{n}}\right) + -3.7 = -4.0 + -3.7 = -7.7$
4. $\delta_2(n) = \max\left(\underbrace{-5.3}_{\text{a}}, \underbrace{-6.9}_{\text{v}}, \underbrace{-2.5}_{\textbf{d}}, \underbrace{-6.9}_{\text{n}}\right) + -1.9 = -2.5 + -1.9 = -4.4$

Example ...

Third position

1. $\delta_3(a) = \max\left(\underbrace{-5.0}_{\mathbf{a}}, \underbrace{-8.6}_{\text{v}}, \underbrace{-8.6}_{\text{d}}, \underbrace{-7.4}_{\text{n}}\right) + -2.3 = -5.0 + -2.3 = -7.3$

Example ...

Third position

1. $\delta_3(a) = \max\left(\underbrace{-5.0}_{\mathbf{a}}, \underbrace{-8.6}_{\text{v}}, \underbrace{-8.6}_{\text{d}}, \underbrace{-7.4}_{\text{n}}\right) + -2.3 = -5.0 + -2.3 = -7.3$
2. $\delta_3(v) = \max\left(\underbrace{-6.1}_{\text{a}}, \underbrace{-8.6}_{\text{v}}, \underbrace{-10.7}_{\text{d}}, \underbrace{-4.6}_{\mathbf{n}}\right) + -0.9 = -4.6 + -0.9 = -5.5$

Example ...

Third position

1. $\delta_3(a) = \max\left(\underbrace{-5.0}_{\mathbf{a}}, \underbrace{-8.6}_{\mathrm{v}}, \underbrace{-8.6}_{\mathrm{d}}, \underbrace{-7.4}_{\mathrm{n}}\right) + -2.3 = -5.0 + -2.3 = -7.3$
2. $\delta_3(v) = \max\left(\underbrace{-6.1}_{\mathrm{a}}, \underbrace{-8.6}_{\mathrm{v}}, \underbrace{-10.7}_{\mathrm{d}}, \underbrace{-4.6}_{\mathbf{n}}\right) + -0.9 = -4.6 + -0.9 = -5.5$
3. $\delta_3(d) = \max\left(\underbrace{-6.1}_{\mathbf{a}}, \underbrace{-8.2}_{\mathrm{v}}, \underbrace{-10.0}_{\mathrm{d}}, \underbrace{-7.4}_{\mathrm{n}}\right) + -3.7 = -6.1 + -3.7 = -9.8$

Example ...

Third position

1. $\delta_3(a) = \max\left(\underbrace{-5.0}_{\mathbf{a}}, \underbrace{-8.6}_{\mathrm{v}}, \underbrace{-8.6}_{\mathrm{d}}, \underbrace{-7.4}_{\mathrm{n}}\right) + -2.3 = -5.0 + -2.3 = -7.3$
2. $\delta_3(v) = \max\left(\underbrace{-6.1}_{\mathrm{a}}, \underbrace{-8.6}_{\mathrm{v}}, \underbrace{-10.7}_{\mathrm{d}}, \underbrace{-4.6}_{\mathbf{n}}\right) + -0.9 = -4.6 + -0.9 = -5.5$
3. $\delta_3(d) = \max\left(\underbrace{-6.1}_{\mathbf{a}}, \underbrace{-8.2}_{\mathrm{v}}, \underbrace{-10.0}_{\mathrm{d}}, \underbrace{-7.4}_{\mathrm{n}}\right) + -3.7 = -6.1 + -3.7 = -9.8$
4. $\delta_3(n) = \max\left(\underbrace{-4.5}_{\mathbf{a}}, \underbrace{-8.2}_{\mathrm{v}}, \underbrace{-8.5}_{\mathrm{d}}, \underbrace{-6.7}_{\mathrm{n}}\right) + -0.9 = -4.5 + -0.9 = -5.4$

Example ...

Fourth position

1. $\delta_4(a) = \max\left(\underbrace{-8.5}_{\text{a}}, \underbrace{-7.2}_{\textbf{v}}, \underbrace{-10.7}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -3.4 = -7.2 + -3.4 = -10.6$

Example . . .

Fourth position

1. $\delta_4(a) = \max\left(\underbrace{-8.5}_{\text{a}}, \underbrace{-7.2}_{\textbf{v}}, \underbrace{-10.7}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -3.4 = -7.2 + -3.4 = -10.6$
2. $\delta_4(v) = \max\left(\underbrace{-9.6}_{\text{a}}, \underbrace{-7.2}_{\text{v}}, \underbrace{-12.8}_{\text{d}}, \underbrace{-5.7}_{\textbf{n}}\right) + -3.4 = -5.7 + -3.4 = -9.1$

Example ...

Fourth position

1. $\delta_4(a) = \max\left(\underbrace{-8.5}_{\text{a}}, \underbrace{-7.2}_{\mathbf{v}}, \underbrace{-10.7}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -3.4 = -7.2 + -3.4 = -10.6$
2. $\delta_4(v) = \max\left(\underbrace{-9.6}_{\text{a}}, \underbrace{-7.2}_{\text{v}}, \underbrace{-12.8}_{\text{d}}, \underbrace{-5.7}_{\mathbf{n}}\right) + -3.4 = -5.7 + -3.4 = -9.1$
3. $\delta_4(d) = \max\left(\underbrace{-9.6}_{\text{a}}, \underbrace{-6.8}_{\mathbf{v}}, \underbrace{-12.1}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -0.5 = -6.8 + -0.5 = -7.3$

Example ...

Fourth position

1. $\delta_4(a) = \max\left(\underbrace{-8.5}_{\text{a}}, \underbrace{-7.2}_{\textbf{v}}, \underbrace{-10.7}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -3.4 = -7.2 + -3.4 = -10.6$
2. $\delta_4(v) = \max\left(\underbrace{-9.6}_{\text{a}}, \underbrace{-7.2}_{\text{v}}, \underbrace{-12.8}_{\text{d}}, \underbrace{-5.7}_{\textbf{n}}\right) + -3.4 = -5.7 + -3.4 = -9.1$
3. $\delta_4(d) = \max\left(\underbrace{-9.6}_{\text{a}}, \underbrace{-6.8}_{\textbf{v}}, \underbrace{-12.1}_{\text{d}}, \underbrace{-8.4}_{\text{n}}\right) + -0.5 = -6.8 + -0.5 = -7.3$
4. $\delta_4(n) = \max\left(\underbrace{-8.0}_{\text{a}}, \underbrace{-6.8}_{\textbf{v}}, \underbrace{-10.6}_{\text{d}}, \underbrace{-7.7}_{\text{n}}\right) + -3.4 = -6.8 + -3.4 = -10.2$

Example ...

Fifth position

1. $\delta_5(a) = \max\left(\underbrace{-11.8}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-8.2}_{\textbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -2.3 = -8.2 + -2.3 = -11$

Example ...

Fifth position

1. $\delta_5(a) = \max\left(\underbrace{-11.8}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-8.2}_{\mathbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -2.3 = -8.2 + -2.3 = -11$

2. $\delta_5(v) = \max\left(\underbrace{-12.9}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-10.3}_{\mathbf{d}}, \underbrace{-10.4}_{\text{n}}\right) + -1.6 = -10.3 + -1.6 =$
   -12

Example ...

Fifth position

1. $\delta_5(a) = \max\left(\underbrace{-11.8}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-8.2}_{\textbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -2.3 = -8.2 + -2.3 = -11$
2. $\delta_5(v) = \max\left(\underbrace{-12.9}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-10.3}_{\textbf{d}}, \underbrace{-10.4}_{\text{n}}\right) + -1.6 = -10.3 + -1.6 =$ -12
3. $\delta_5(d) = \max\left(\underbrace{-12.9}_{\text{a}}, \underbrace{-10.3}_{\text{v}}, \underbrace{-9.6}_{\textbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -3.7 = -9.6 + -3.7 = -13$

Example . . .

Fifth position

1. $\delta_5(a) = \max\left(\underbrace{-11.8}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-8.2}_{\textbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -2.3 = -8.2 + -2.3 = -11$
2. $\delta_5(v) = \max\left(\underbrace{-12.9}_{\text{a}}, \underbrace{-10.7}_{\text{v}}, \underbrace{-10.3}_{\textbf{d}}, \underbrace{-10.4}_{\text{n}}\right) + -1.6 = -10.3 + -1.6 =$
   -12
3. $\delta_5(d) = \max\left(\underbrace{-12.9}_{\text{a}}, \underbrace{-10.3}_{\text{v}}, \underbrace{-9.6}_{\textbf{d}}, \underbrace{-13.2}_{\text{n}}\right) + -3.7 = -9.6 + -3.7 = -13$
4. $\delta_5(n) = \max\left(\underbrace{-11.3}_{\text{a}}, \underbrace{-10.3}_{\text{v}}, \underbrace{-8.1}_{\textbf{d}}, \underbrace{-12.5}_{\text{n}}\right) + -1.2 = -8.1 + -1.2 = -9.3$

Example . . .

Reconstruction

Example . . .

Reconstruction

For “the old man”, the reconstruction starts with the best part of speech at Position 3, which is noun (-5.4), which has an adjective back pointer, which as a back pointer to determiner. The overall sequence is “The/det old/adj man/n”.

Example . . .

Reconstruction

For “the old man”, the reconstruction starts with the best part of speech at Position 3, which is noun (-5.4), which has an adjective back pointer, which as a back pointer to determiner. The overall sequence is “The/det old/adj man/n”.

For “the old man the boats”, the reconstruction starts with the best part of speech at Position 5, which is a noun (-9.3), which leads to the sequence “The/det old/n man/v the/det boats/n”.