Hide menu

Lecture: Part of Speech Tagging

- **Video:** Week Introduction 1 min
- **Video:** Part of Speech Tagging 2 min
- **Reading:** Part of Speech Tagging 4 min
- **Lab:** Lecture Notebook - Working with text files 20 min
- **Video:** Markov Chains 3 min
- **Reading:** Markov Chains 3 min
- **Video:** Markov Chains and POS Tags 4 min
- **Reading:** Markov Chains and POS Tags 6 min
- **Video:** Hidden Markov Models 3 min
- **Reading:** Hidden Markov Models 6 min
- **Video:** Calculating Probabilities 3 min
- **Reading:** Calculating Probabilities 5 min
- **Video:** Populating the Transition Matrix 4 min
- **Reading:** Populating the Transition Matrix 6 min
- **Video:** Populating the Emission Matrix 2 min
- **Reading:** Populating the Emission Matrix 5 min
- **Lab:** Lecture Notebook - Working with tags and Numpy 20 min
- **Video:** The Viterbi Algorithm 4 min
- **Reading:** The Viterbi Algorithm 5 min
- **Video:** Viterbi: Initialization 2 min
- **Reading:** Viterbi Initialization 5 min
- **Video:** Viterbi: Forward Pass 2 min
- **Reading:** Viterbi: Forward Pass 10 min
- **Video:** Viterbi: Backward Pass 5 min
- **Reading:** Viterbi: Backward Pass 10 min

Week 2 > Populating the Transition Matrix

Previous Next

Populating the Transition Matrix

To populate the transition matrix you have to keep track of the number of times each tag shows up before another tag.

$A =$

	NN	VB	O
π	1	0	2
NN (noun)	0	0	6
VB (verb)	0	0	0
O (other)	6	0	8

<s> in a station of the metro
<s> the apparition of these faces in the crowd :
<s> petals on a wet , black bough .

Ezra Pound – 1913

In the table above, you can see that green corresponds to nouns (NN), purple corresponds to verbs (VB), and blue corresponds to other (O). Orange (π) corresponds to the initial state. The numbers inside the matrix correspond to the number of times a part of speech tag shows up right after another one.

To go from O to NN or in other words to calculate $P(O|NN)$ you have to compute the following:

$A =$

	NN	VB	O	
π	1	0	2	3
NN	0	0	6	6
VB	0	0	0	0
O	6	0	8	14

$$P(\text{NN}|\text{O}) = \frac{C(\text{O}, \text{NN})}{\sum_{j=1}^{N} C(\text{O}, t_j)} = \frac{6}{14}$$

To generalize:

$$P(t_i \mid t_{i-1}) = \frac{C(t_{i-1}, t_i)}{\sum_{j=1}^{N} C(t_{i-1}, t_j)}$$

Unfortunately, sometimes you might not see two POS tags in front each other. This will give you a probability of 0. To solve this issue, you will "smooth" it as follows:

$A =$

	NN	VB	O	
π	1+ε	0+ε	2+ε	3+3*ε
NN	0+ε	0+ε	6+ε	6+3*ε
VB	0+ε	0+ε	0+ε	0+3*ε
O	6+ε	0+ε	8+ε	14+3*ε

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i) + \epsilon}{\sum_{j=1}^{N} C(t_{i-1}, t_j) + N * \epsilon}$$

The ϵ allows you to not have any two sequences showing up with 0 probability. Why is this important?

Mark as completed

Like Dislike Report an issue