



Natural Language

Practical Classes

Luísa Coheur

2025

P7

Morphology



Image generated by ChatGPT

- **Summary:**

- The Viterbi algorithm
- POS tagging

- **Operational objectives:**

- Practice the Viterbi algorithm
- Play with NLP tools (NLTK and SpaCy)

- **This class needs:** paper, a pen/pencil and computer

- **Class material:** these guidelines, a notebook and a test file

I. Viterbi or your first client of the day

Inspector Morcela, contacts you again:

Youngster,

There is this NLP freak who stole several items from the police squad. Now he is threatening the police, saying he is going to destroy (with a bomb) a hard disk with very important information, unless we give him 100k €. He likes to play games and said that he will send us a quiz. He mentioned someone named Viterbi. Might be an accomplice. Either way, stay alert, we will need your expertise and will send you the quiz as soon as it comes into our hands.

Viterbi? That name rings a bell! You decide to study the Viterbi algorithm. You need to be ready to help Morcela.

You check the Viterbi algorithm, as described in Figure 1:

```

 $i \leftarrow 1$ 
while  $i < N$  do
     $SS(i, 1) = P(w_1 | L_i) * P(L_i | < s >)$ 
     $BP(i, 1) = 0$ 
     $i++$ 
end while
 $t \leftarrow 2$ 
while  $t < n$  do
     $i \leftarrow 1$ 
    while  $i < N$  do
         $SS(i, t) = \max_{j=1, \dots, N} SS(j, t-1) * P(L_j | L_i) * P(w_t | L_i)$ 
         $BP(i, t) = j$  that resulted in the maximum score
         $i++$ 
    end while
     $t++$ 
end while
 $C(n) = i$  that maximizes  $SS(i, n)$ 
 $i \leftarrow n - 1$ 
while  $i > 1$  do
     $i--$ 
     $C(n) = BP(C(i+1), i+1)$ 
end while

```

- N = number of tags
- n = number of words in the sequence
- Data structures:
 - SS (Sequence score) – records the score of the best sequence found up to a given position with category L .
 - BP (Back Pointer) – records the previous state to a given state
 - C – records the best sequence of tags.

Figura 1: The Viterbi algorithm

1. You consider the following probabilities (assume all non-reported ones are 0):

- $P(\text{will} | \text{Noun}) = 0.2$; $P(\text{will} | \text{Verb}) = 0.25$
- $P(\text{cook} | \text{Noun}) = 0.2$; $P(\text{cook} | \text{Verb}) = 0.1$
- $P(\text{Noun} | <s>) = 0.7$; $P(\text{Noun} | \text{Verb}) = 0.4$; $P(\text{Noun} | \text{Noun}) = 0.2$
- $P(\text{Verb} | <s>) = 0.3$; $P(\text{Verb} | \text{Verb}) = 0.55$; $P(\text{Verb} | \text{Noun}) = 0.4$

- (a) You consider the lowercased sentence *will cook will* and apply Viterbi to find which is the most probable sequence of PoS tags for that sentence.
- (b) Now consider the sequence *will cook*. Without further counts, what is the most probable sequence of labels? (Dynamic programming is so cool!)

2. Finally, you receive the quiz that the NLP freak sent to Morcela:

Despicable Morcela,

You will never be able to get this: the code is the best sequence for "YOU ARE SO STUPID". I'm sure you will fail. Muahahahahah!

Table SS	YOU	ARE	SO	STUPID
1	0.3	0.01	0.001	0.0001
2	0.4	0.015	0.005	0.0002
3	0.02	0.02	0.0045	0.0001
4	0.1	0.01	0.00887	0.00013
Table BP	YOU	ARE	SO	STUPID
1	0	4	3	2
2	0	1	2	1
3	0	4	3	4
4	0	1	1	3

Tabela 1: NLP freak quiz. Note that values can be unreal.

You analyse the table the NLP freak sent:

You know what to do! You are the GOAT! After a while, you call Inspector Morcela.

(a) What is the code?

Inspector Morcela finally replied:

– Ok, I trust you. I will introduce it now.

Silence.

More silence.

No BOOM!!!!

You start to feel hopeful that it is done, but then, suddenly, Morcela says:

– Oh, the code was correct, but there is another code and another letter. It says:

Oopsy! New sentence: “YOU ARE”. You have 1 minute or... BOOM!!!

Ok, I can get this!

(b) Find the second code, hopefully in 20 seconds.

No BOOM again!!!! Is this nightmare over?

II. Morphology or your second client of the day

Your heart still races with adrenaline when you receive a message from another client. Your next client is a 9-year-old boy:

Dear Sir,

Thank you so much for your kind help! My teacher gave us a really interesting assignment about Little Red Cap. We have two texts: one has lots of adjectives, and the other one is full of adverbs. We need to figure out which one has the adverbs and which one has the adjectives. Then, we need to find all the adverbs and adjectives in each text and count them. I think I know the answer, but I would be so very grateful if you could help me check my work. I promise I will give you my allowance if you help me with the solution so I can make sure mine is correct! I will send you the file in a few minutes.

Thank you again, Sir!

James Fitzgerald Rockefeller IV

Hands on

Well, perfect timing! You are studying morphology in your Natural Language classes. Besides, so polite and with such a name, James probably has a very nice allowance. Before he sends the texts, you decide to practice a little bit.

1. Practicing Morphology

Fill in the table by associating each form of word construction with one of the following words: *cat, criminals, unhappy, shamelessness*.

Word Formation Method	Word
Derivation	
Inflection	
Compounding (agglutination)	
None of the above	

Tabela 2: Word Formation Table

2. Taking advantage of widely used NLP tools

Finally, James has sent you the file with the texts:

- Text A:

In a dense, shadowy forest lived a charming, young girl named Little Red Cap. She wore a vibrant, scarlet cloak that fluttered with every graceful step she took. The path she followed was winding and narrow, surrounded by towering, ancient trees with gnarled branches. Her delicate basket was filled with freshly baked, fragrant goodies for her sweet, elderly grandmother. The day was bright, the sky painted a brilliant azure with fluffy white clouds drifting lazily. Little Red Cap's heart was pure and her spirit fearless, unaware of the cunning, hungry wolf lurking in the dark woods. The wolf had sharp, menacing eyes and a sly, wicked grin that could send shivers down anyone's spine. He was a creature of immense deception, always on the lookout for his next unsuspecting prey. As she neared her grandmother's cozy, quaint cottage, the air became thick with an eerie, unsettling silence.

- Text B:

Little Red Cap walked happily through the forest, holding her basket tightly. She smiled warmly as the birds chirped cheerfully above her head. The sun shone brightly, casting its rays gently on the path. She moved quickly, eager to reach her grandmother's house before it got too late. The wolf, however, watched her silently from behind the trees, creeping closer slowly. He spoke sweetly when he approached her, deceiving her cleverly with his words. Unaware of his intentions, she listened politely, responding trustfully to his questions. The wolf, meanwhile, plotted deviously, planning to reach her grandmother's house swiftly. When Little Red Cap arrived, she knocked gently on the door, waiting patiently for a response.

Now it is time to take the case of young James into our hands.

You decide to test a jupyter notebook that uses the NLP tools NLTK and SpaCy. You run it (use Google Colab¹ and don't forget to upload the given auxiliary file).

You adapt it to solve James case. Which one is more rich in adverbs? And in adjectives?

Nice. You did it!

You reply to James with your response.

After a few minutes, he answers back to you: he thanks you a lot and then asks for your bank account details, saying his butler will transfer his allowance to you. His allowance is only \$2 a month, and he explains that his parents believe children should learn the value of money.

You can't help but sigh. Huff. Will you ever be rich?

¹<https://colab.research.google.com>