

CSCI.GA.2590 - Natural Language Processing – Spring 2016 – Prof. Grishman

Assignment #3

1. HMM

- (a) (2 points) Consider a HMM with two states, Cow and Duck, and a start and end state. Emission probabilities:
- In state Cow, the HMM can emit 'moo' (with 0.9 probability) or 'hello' (0.1 probability).
 - In state Duck, the HMM can emit 'quack' (0.6 probability) or 'hello' (0.4 probability). The Duck has been studying English longer.

(Nothing is emitted in the start or end state.) Transition probabilities:

- From the start state, the HMM goes to state Cow with 1.0 probability (i.e., always).
- From state Cow, the HMM can remain in state Cow (0.5 probability), go to state Duck (0.3 probability), or go to state end (0.2 probability).
- From state Duck, the HMM can remain in state Duck (0.5 probability), go to state Cow (0.3 probability), or go to state end (0.2 probability).

Using the Viterbi algorithm, decode (find the most likely state sequence for) 'moo hello quack'. What is the probability of emitting this sentence from this state sequence? Show your work, so that you can get partial credit even if you make an error.

- (b) (1 point) Is there another state sequence which also generates 'moo hello quack'? What is the total probability of emitting this sentence?

Solution

- (a) Transition and emission probabilities are shown below:

$$\begin{aligned}
 a(Cow|S) &= 1, a(Duck|S) = 0, \\
 a(Cow|Cow) &= 0.5, a(Duck|Cow) = 0.3, a(E|Cow) = 0.2, \\
 a(Duck|Duck) &= 0.3, a(Cow|Duck) = 0.3, a(E|Duck) = 0.2 \\
 b("moo"|Cow) &= 0.9, b("hello"|Cow) = 0.1, \\
 b("quack"|Duck) &= 0.6, b("hello"|Duck) = 0.4,
 \end{aligned}$$

DP table is shown below:

Start	moo	hello	quack	End
Cow	$a(Cow S) \cdot b("moo" Cow) = 0.9$	$\max(0.9 \times 0.5 \times 0.1, 0) = 0.045[Cow]$	0	0
Duck	0	$\max(0.9 \times 0.3 \times 0.4, 0) = 0.108[Cow]$	$0.0324[Duck]$	0.00648

From End state, we backtrack to Start state, so the most likely state sequence is **Cow - Duck - Duck**, whose total probability is 0.00648.

- (b) Another sequence that generates "moo hello quack" is **Cow - Cow - Duck**, with probability 0.00162.

2. JET HMM Tagger

Solution

In second example, state VBD emitting “said” has 6879 instances, while state VBN emitting “said” has 88, which is almost two magnitude smaller. So the likelihood of being VBD is much larger than being VBN. But the correct tag should be VBN, the past participle.

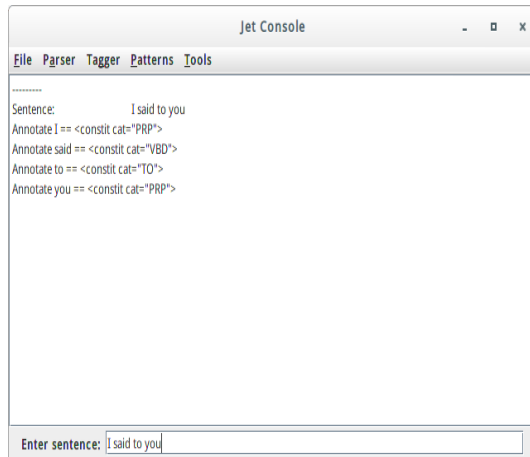


Figure 1: Correctly tagged - “I said to you.”

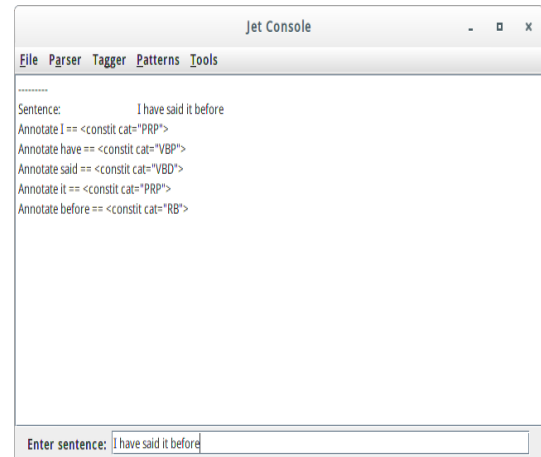


Figure 2: One incorrectly tagged - “I have said it before.”