

Natural Language Processing (JHU 601.465/665)  
 Answers to "HMMs and EM" practice problems

1. (a) Part of speech tagging.
- (b) Viterbi algorithm.
- (c) states = tags  
 $\text{days} = \text{words}$   
 $\text{dollars} = \text{costs}$  (negative log probabilities)
- (d) A state of the trellis corresponds to being at a particular festival on a particular day. Let  $\mu(y)$  represent the worth of the optimal path to state  $y$  in the trellis.

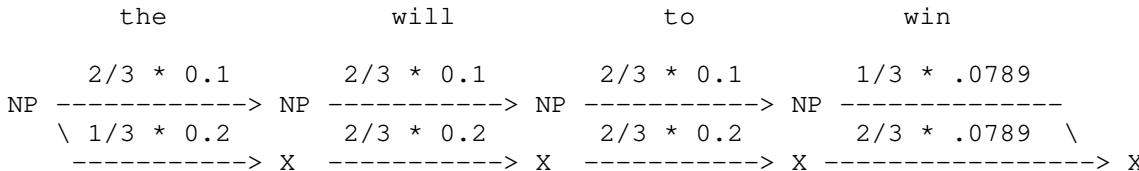
For each state  $y$ , you must compute

$$\begin{aligned}\mu(y) = & \text{ (maximum over all arcs } x \rightarrow y \\ & \text{ of } (\mu(x) - \text{cost of getting from } x \text{ to } y)) \\ & + \text{benefit of being in state } y\end{aligned}$$

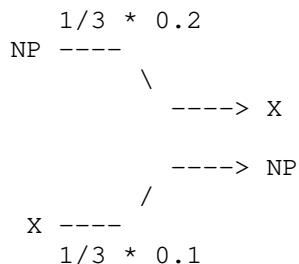
So each state  $y$  requires an addition ("maximum ... + benefit") and each arc  $x \rightarrow y$  requires a subtraction (" $\mu(x) - \text{cost ...}$ "). The trellis has  $50 \times 30$  states, since you can be in any of 50 places on each of the next 30 days. It has  $1 \times 50 + 50 \times 50 \times 29$  arcs, since you can go from Baltimore to anywhere on the first day, and you can go from anywhere to anywhere on the remaining 29 days. So the total number of additions is  $50 \times 30 + 1 \times 50 + 50 \times 50 \times 29 = 74,050$ .

2. (a) "will." It gets the most NP influence from "the." As we move rightward, this influence dies down and we start getting more X influence from the period.

- (b) The probability works out to  $1/4$ . The trellis looks like this:



both "will" and "win" also have diagonal arcs not shown above:



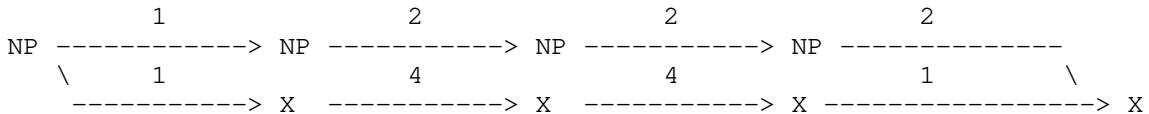
It's easy to find the alpha and beta probabilities at the word "to." Each one is a sum over only two paths:

$$\begin{aligned}\alpha(\text{to}/\text{NP}) &= 6/900 \quad \beta(\text{to}/\text{NP}) = 6/90 * .0789 \\ \alpha(\text{to}/\text{X}) &= 12/900 \quad \beta(\text{to}/\text{X}) = 9/90 * .0789\end{aligned}$$

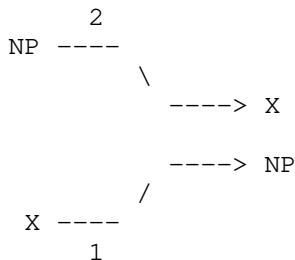
It follows that the total probability of all paths through to/NP is  $\alpha(\text{to}/\text{NP}) * \beta(\text{to}/\text{NP}) = 6*6/(900*90) * .0789$  and the total probability of all paths through X at "to" is  $\alpha(\text{to}/\text{X}) * \beta(\text{to}/\text{X}) = 12*9/(900*90) * .0789$

so the paths through to/NP have  $6*6 / (6*6+12*9) = 1/4$  of the total probability mass.

If you took the hint, you would have considered only the relative probabilities at each time step, since that's all that really matters. This makes the arithmetic much easier:



again, both of the middle two sections also have diagonal arcs not shown above:



Now we get

$$\begin{array}{ll} \text{alpha(NP)} = 3 & \text{beta(NP)} = 6 \\ \text{alpha(X)} = 6 & \text{beta(X)} = 9 \end{array}$$

and it is much easier to see that the paths through NP have 1/4 of the total probability mass.

If we considered the whole sentence, including "Popeye has," it wouldn't change the answer. Multiplying in "Popeye has" would have identical effect on the paths through to/NP and the paths through to/X, since both sets of paths have to go through the/NP anyway. The unambiguous "the" and "." act like sentence boundaries, insulating the ambiguous parts of the input from one another.

- 3. a. 25. Removing the three "right parse only" rules (probability  $0.1 * 0.1 * 0.03$ ) and replacing them with the three "left parse only" rules (probability  $0.5 * 0.5 * 0.03$ ) increases the probability by a factor of 25.
- b. The rules that appear in both parses have probability 1. The rules that appear in only the left parse have probability  $25/26$ , since the left parse is 25 times as likely. The rules that appear in only the right parse have probability  $1/26$ .
- c. The grammar's different competing ways of rewriting NP must have probabilities that sum to 1.

In the training corpus (i.e., this training sentence), we just saw that the fractional counts of these 4 NP rules are  $25/26$ ,  $1/26$ , 1, and  $1/26$ . So NP was expected to be observed  $25/26 + 1/26 + 1 + 1/26 = 53/26$  times in total.

Thus, the 4 rewrites were respectively  $25/53$ ,  $1/53$ ,  $26/53$ , and  $1/53$  of the total. These are the new probabilities.

- d. There is only one way of getting from NP to "her coins." This involves the left-parse rules
  - $0.5 \text{ NP} \rightarrow \text{Det N}$
  - $0.03 \text{ Det} \rightarrow \text{her}$
  - $0.001 \text{ N} \rightarrow \text{coins}$

The product of their probabilities is  $1.5e-5$ , the inside probability.

There is also only one way of getting from S to "Everyone threw NP." This involves all the other left-parse rules:

0.8 S -> NP VP  
0.001 NP -> Everyone  
0.5 VP -> V NP  
0.001 V -> threw.

The product of their probabilities is 4e-7, the outside probability.

(If there were any other ways of doing these derivations, for example using other grammar rules not shown in the question, then parser would have found additional parses. In that case, we would have had to sum the probabilities of multiple derivations to get the inside or outside probabilities.)

- e. It's not in Chomsky Normal Form.  
(Specifically, it has the rules NP -> N and VP -> V NP NP.)
- f. O(n^3).
- g. NP can start with Det, N, Everyone, or her.  
Det can start with her.  
N can start with coins.

So the possible left corners of NP are  
{Det,N,Everyone,her,coins}: these are the words or nonterminals that can start an NP.

We also gave credit for the answer {Everyone,her,coins}.