

050/520/600.666 Information Extraction from Speech and Text

Homework # 2

Due February 17, 2006.

1. Consider the HMM of Chapter 2, Figure 2.8, with state space $\mathcal{S} = \{1, 2, 3\}$ and output alphabet $\mathcal{Y} = \{0, 1\}$. Using the notation of Chapter 2, let the *transition* probabilities for the output-producing transitions and null transitions be given, respectively, by

$$p(s'|s) \equiv \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{3} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{bmatrix} \quad \text{and} \quad q(s'|s) \equiv \begin{bmatrix} 0 & 0 & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}, \quad s, s' \in \mathcal{S}.$$

Let the probability of emitting a “0” during an output producing transition from state s to state s' be

$$q(0|s \rightarrow s') \equiv \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}, \quad s, s' \in \mathcal{S},$$

and note that $q(1|s \rightarrow s')$ is simply $1 - q(0|s \rightarrow s')$. Let the initial state be $s_0 = 1$.

Perform the following calculations *by hand*, retaining intermediate answers in fractional form to preserve numerical accuracy.

- (a) Draw a state diagram of this HMM, attaching state-labels to all nodes, probabilities $r(y, s'|s)$ or $q(s'|s)$ to all arcs and output-labels to all non-null arcs.
- (b) Draw a 4-stage trellis for this HMM, showing *only* the paths which could have resulted in the output 0110.
- (c) Compute the *forward* probabilities $\alpha_i(s)$ for the output sequence 0110, and indicate them on the trellis drawn above for (b).
- (d) Compute the marginal probability of the output: $P(y_1y_2y_3y_4 = 0110|s_0 = 1)$.
- (e) Redraw the trellis with paths which could have given rise to 0110, and indicate on it the *Viterbi* probabilities $\gamma_i(s)$. Color the most like path given $y_1y_2y_3y_4 = 0110$.
- (f) Redraw the trellis, showing paths which could have given rise to 0110, and compute the *backward* probabilities $\beta_i(s)$. Indicate them on this redrawn trellis.
- (g) Compare $\beta_0(1)$ with the marginal probability $P(y_1y_2y_3y_4 = 0110|s_0 = 1)$ of (d).

- (h) Calculate the *a posteriori* probabilities $P(t^i = t \mid y_1 y_2 y_3 y_4 = 0110, s_0 = 1)$ for each arc in the trellis of (f). Show your answers on the trellis.
- (i) Based on your calculations in (h), compute the *expected* counts $c(t)$ of each arc, and *reestimate* the HMM transition probability matrices $p(s'|s)$ and $q(s'|s)$.
- Caution: the matrix $p(s'|s)$ has only one entry for a non-null transition from s to s' , while our trellis has up to $|\mathcal{Y}|$ distinct non-null arcs from s to s' , one arc per output symbol. Make sure you consolidate their counts appropriately.
- (j) Based on your calculations in (h), compute the *expected* counts $c(y, t)$ of each non-null arc, and *reestimate* the HMM emission probability matrix $q(0|s \rightarrow s')$.
2. *Eliminating null arcs:* Compute the probabilities $v(s'|s)$ of going from s to s' via any number of null arcs, including zero null arcs when $s = s'$, and use them to construct a new HMM with no null arcs that is *equivalent* to the original HMM of Problem 1. Write down the matrices $p(s'|s)$ and $q(0|s \rightarrow s')$ for this HMM, and draw its state transition diagram.
3. *Outputs on States:* Construct an HMM which produces outputs from states, instead of arcs, that is *equivalent* to the original HMM of Problem 1. In particular, construct an equivalent HMM with the minimum number of states, and specify the transition probability matrix $p(s'|s)$, and output probabilities $q(0|s)$ for each state.
4. *Two Best Paths:* Devise a two-pass algorithm to find the *two most likely paths* through the HMM trellis. Specifically, let the first pass be the Viterbi algorithm, and devise a second pass over the same trellis to find the second most likely path *knowing* the most likely path. Assume if needed that there are no cycles made up entirely of null arcs.
- (a) Describe your second pass in the same way the Viterbi algorithm is described in Chapter 2.
- (b) Redraw the trellis of Problem 1(e), run your algorithm on it by hand, and color the two most likely paths.