050/520/600.666 Information Extraction from Speech and Text

Homework # 2

Due February 18, 2005.

1. Write a **two page** summary (including any figures if you wish) of the article

   L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, pp 4-16, Jan 1986.

2. Consider an HMM 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

$$r(j|i) \equiv \begin{bmatrix} \frac{2}{3} & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{bmatrix} \quad \text{and} \quad q(j|i) \equiv \begin{bmatrix} 0 & 0 & \frac{1}{6} \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}, \quad i, j \in \mathcal{S}.$$

   Let the probabilitity of emitting a "0" during an output producing transition from state i to state j be

$$w(0|i \to j) \equiv \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}, \quad i, j \in \mathcal{S},$$

   and note that $w(1|i \to j)$ is simply $1 - w(0|i \to j)$.

   (a) Draw a state diagram of this HMM, attaching state-labels to all nodes, probabilities to all arcs and output-labels to all output-producing arcs (cf Figure 2.8).

   (b) Given $s_0 = 1$, and the output sequence 0110, draw a 4-stage trellis for this HMM showing only the paths which could have resulted in this output.

   (c) Construct an HMM which produces outputs from states that is *equivalent* to the HMM defined above. In particular, construct an equivalent HMM with the minimum number of states, and specify the transition and output probabilities for each state.

3. *HMMs with State Durations*: Consider an HMM with outputs produced by states. Let $\mathcal{S}$ denote the state-space and $\mathcal{Y}$ the discrete and finite output alphabet, as usual, and let the transition and output probabilities be denoted by

$$a_{ij} = P(s_t = j | s_{t-1} = i) \quad \text{and} \quad b_j(k) = P(y_t = k | s_t = j) \quad i, j \in \mathcal{S}, \quad k \in \mathcal{Y}.$$

Recall that when $a_{ii} > 0$, the underlying Markov chain may *loop around* in the state i for more than one time-step.

   (a) Show that the *state durations* of the underlying Markov chain have a geometric distribution. Specifically, argue why for $t \geq 0$,

$$P(s_{t+1} = \ldots = s_{t+\tau-1} = j,\ s_{t+\tau} \neq j | s_t = j) \;=\; (1 - a_{jj}) a_{jj}^{\tau-1}, \quad \tau = 1, 2, \ldots.$$

   (b) Does the probability calculated above depend on how long prior to t the Markov chain has already been in state j?

Consider *modifying* the HMM to explicitly model state durations. Given the probability distributions $\delta_i(\tau)$ of the duration for each state i, where $\tau = 1, 2, \ldots, T$ and $\sum_{\tau=1}^{T} \delta_i(\tau) = 1$, an initial state s_0, and $a_{ii} = 0$, the (hidden) state process

- *decides* to stay in the initial state s_0 for a duration τ with probability $\delta_{s_0}(\tau)$,
- *produces* the first τ outputs $y_1, \ldots, y_\tau$ from s_0 with probability $\prod_{t=1}^{\tau} b_{s_0}(y_t)$,
- *makes a transition* to reach state $s_{\tau+1} \neq s_0$ with probability $a_{s_0 s_{\tau+1}}$,
- *decides* to stay in the state $s_{\tau+1}$ for a duration τ' with probability $\delta_{s_{\tau+1}}(\tau')$,
- *produces* the next τ' outputs $y_{\tau+1}, \ldots, y_{\tau+\tau'}$ from state $s_{\tau+1}$ with probability $\prod_{t=1}^{\tau'} b_{s_{\tau+1}}(y_{\tau+t})$,
- *makes a transition* to reach state $s_{\tau+\tau'+1} \neq s_{\tau+1}$ with probability $a_{s_{\tau+1} s_{\tau+\tau'+1}}$,

and so on. This model appears to be richer at first glance than an HMM.

   (c) Show that the model described above is equivalent to an HMM. Draw a simple example with $|\mathcal{S}| = 2$ and $T = 2$.

   (d) If $\alpha_t(j)$ is the probability that the state process has *just reached* the state j after having produced $y_1, \ldots, y_t$, compute the *forward probability* $\alpha_t(j)$ in terms of $\alpha_{t'}(i)$, $t' < t$.

4. 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 transitions. Describe your second pass in the same way the Viterbi algorithm is described in Chapter 2.