(Ungraded) Pop Quiz III

CS 753 Automatic Speech Recognition

October 5, 2017

Name: _____ (Optional)

1 Understanding Human Emotions

1. Given an audio stream of a person talking, we would like to model how their emotions change over time and use this information to further predict their personality type. For each speech frame, we extract acoustic features that determine the underlying intonation patterns in speech. Let the intonation pattern at time t be denoted by C_t where $C_t \in \{$flat, rising, falling, smooth, jagged$\}$. We would like to predict the speaker's emotion at each time frame, E_t, whose domain is $\{$disgust, joy, sadness, fear, anger, surprise$\}$.

(A) We want to set up this task as a first-order HMM where E_t denotes the hidden variable and C_t denotes the observation at time t. Describe the number of observation probabilities and the number of transition probabilities that will need to be estimated.

Ans:

(B) Say that we want your HMM defined in part (A) to be a second-order HMM (i.e. the state at time $t + 1$ depends on the states at both time step t and $t - 1$) and we want to derive the new forward algorithm update. Complete the following expression:

$$\Pr(E_t, E_{t+1}|C_{1:t}) \propto \sum_{E_{t-1}} \Pr(E_{t-1}, E_t, C_{1:t-1}) \cdot \underline{\hspace{4cm}}$$

For your reference, below we show how the recursive update in the forward algorithm can be derived for a first-order HMM (where h_t denotes the hidden state at time t, O_t is the observation at time t and $O_{1:t}$ denotes all the observations $O_1, \ldots, O_t$).

$$\Pr(h_t|O_{1:t}) \propto \Pr(h_t, O_{1:t})$$
$$= \sum_{h_{t-1}} \Pr(h_{t-1}, h_t, O_{1:t})$$
$$= \sum_{h_{t-1}} \Pr(h_{t-1}, h_t, O_{1:t-1}, O_t)$$
$$= \Pr(O_t|h_t) \sum_{h_{t-1}} \Pr(h_t|h_{t-1}) \Pr(h_{t-1}, O_{1:t-1})$$

(C) The state sequence from your HMM will correspond to a sequence of emotions. Our final goal is to categorize people into personality types depending on the sequence of emotions they exhibit. Say we define three personality types, "Has anger issues", "Happy-go-lucky person", "Stressed in life". A person who exhibits the emotion "anger" in three or more consecutive time steps is of type "Has anger issues." A "Happy-go-lucky person" never displays the emotions "anger", "fear" and "sadness". A sequence of only "fear" and "sadness" denotes someone who is "Stressed in life". Draw an FST that encodes these constraints and only accepts a sequence of emotions that maps to one of these three personality types.