CSME 206A Natural language & Speech Processing **Spring Semester**

Lecture 8: Hidden Markov Model for Speech Recognition

Lecturer: K.R. Chowdhary *: Professor of CS*

8.1 Weighted Automata v/s HMM (Hidden MArkov Model)

The weighted automata is also called Markov chains. Here, the automata consists of a sequence of states $q = (q_0 q_1 \dots q_n)$, each corresponding to a phone, and set of transition probabilities between states, a_{01}, a_{12}, a_{13}, encodes the probability of one phone following to other (see figure 8.1). To compute the transition probability of a sequence of phones $O = (o_1 o_2 \dots o_t)$, we make use of *forward algorithm*. The figure 8.1 shows the Markov chain for the word "need". The figure shows the transition probabilities and a sample observation sequence. The probabilities are 1 unless otherwise specified.

The figure 8.1 shows the speech input as sequence of symbols. However, in real world, the speech is sliced, ambiguous, real valued input, called *features* or *spectral features*. The second simplification in the Markov chain model above is that, when we see the input symbols [b], we move into the state [b]. In HMM, we cannot look at the input symbols and know which state to move into. In fact, the input symbols do not uniquely determine the next state. For the weighted Automata or Markov chains (called simple Markov model), we use a set of *observation likely-hood B*. The probability $b_i(o_t)$ is 1 if state i is matched to the observation (i.e., phone) o_t and 0 if they did not match.

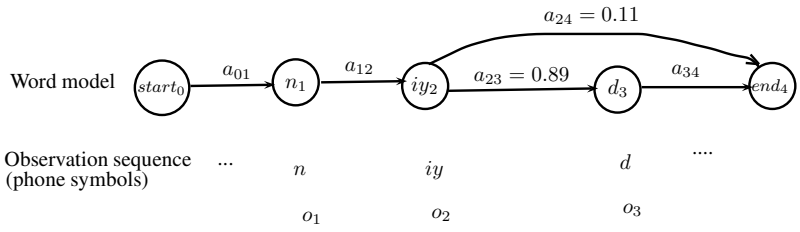

Figure 8.1: Weighted Automata or Markov chain.

An HMM formally differs from a Markov model by adding two more requirements. First, it has a separate set of observable symbols O, which is not drawn from the same alphabet as the state set Q, second, the observation likelihood function B is not limited to the values 1 and 0 (figure 8.1 shows n_1 and iy_2 both as 1). In an HMM the probability $b_i(o_t)$ can take any value from 0 to 1.

The figure 8.2 shows the HMM for the word *need* and sample observation sequence. The obervation sequences are now vectors of spectral features representing the speech signals. Also, we have allowed one state to generate the multiple copies of the same observation, by having a loop on that state. This loop allows HMM to model the variable duration of phones; longer phones requires more loops through the HMM.

In summary, we have following parameters for the HMM:

- States: $Q = q_1 q_2 \dots q_N$, each representing one or more phones,

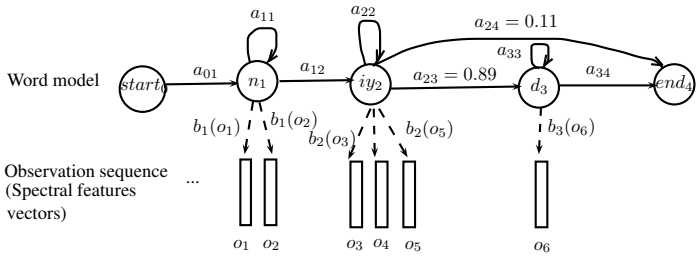

Figure 8.2: An HMM Pronunciation Network for the word *need*.

- Transition probabilities: $A = a_{01}a_{02}\ldots a_{n1}\ldots a_{nn}$. Each a_{ij} represents the probability of transitioning from state i to state j. It is a matrix.
- Observation likelihood: It is a set of observation likelihood $B = b_i(o_t)$, each expressing the probability of an observation o_t being generated from state i.

References

[1] D. Jurafsky and J. Martin, "Speech and Language Processing," *Pearson India*, 2002.