29-03-notes Speech Recognition and Hidden Markov Model

Discrete Markov Process [first order]

The process is modelled in terms of states.

S1,S2,S3,…SN

This process is called first order discrete Markov process if the probability of occurrence of the next state depends only on the current state.

Actual state at time t be qt then

$$P[q_t = S_j | q_{t-1} = S_i]$$

The probability of going from i th to j th state is called transition probability and is given by

$$a_{ij} = P[q_t = S_i | q_{t-1} = S_j] \qquad 1 \leq i, j \leq N$$

The sum of probabilities associated with a state is one because we must either move out of the state or be inside it.

Hidden Markov Model

In our model, the states are hidden from us. We observe some other parameters and predict the probability of the states

We are given the following details

1. The (hidden) states
2. Observables
3. Transition probability: between hidden states
4. Emission probability: probability of occurrences of the observables
5. Initial state distribution

Speech Recognition and HMM

Speech recognition consists of two modules feature extraction and feature matching.

Feature extraction converts speech waveform to some type of representation called feature vector.

In feature matching the feature vector is scored against an acoustic model, the model with max score wins.

Acoustic model representation

Phoneme : minimal unit in a word

CAT -> K A T

Feature vector extracted over 10 to 25 ms is single phoneme. We model this as HMM, each basic unit(phoneme) is given a unique HMM. Each phoneme HMM consists of 3 states, begin middle end. Each state corresponds to a feature vector.

Each word in vocabulary has distinct HMM. When unknown word comes, it is scored against all HMM model and HMM with maximum score is considered as recognized word. By looking dictionary in reverse way (phoneme - word) we can find corresponding word.