Speech Recognition (DSAI 456)

Lecture 5

Mohamed Ghalwash
mghalwash@zewailcity.edu.eg

Lecture 4 Recap

- GMM: Mixture of k Gaussians
- EM algorithm
  - Expectation: Estimate the responsibilities
  - Maximization: Estimate the parameters of the k components
- Speaker Identification: one GMM for each speaker

Agenda

- Motivation
- Mathematical formulation
- Fundamental Problems for HMM
- Forward algorithm
- Using HMMs in speech recognition

Motivation

Source: The actual phonemes or spoken words being uttered by a speaker

Observed: while the acoustic signal features or audio measurements recorded by a microphone.

Motivation

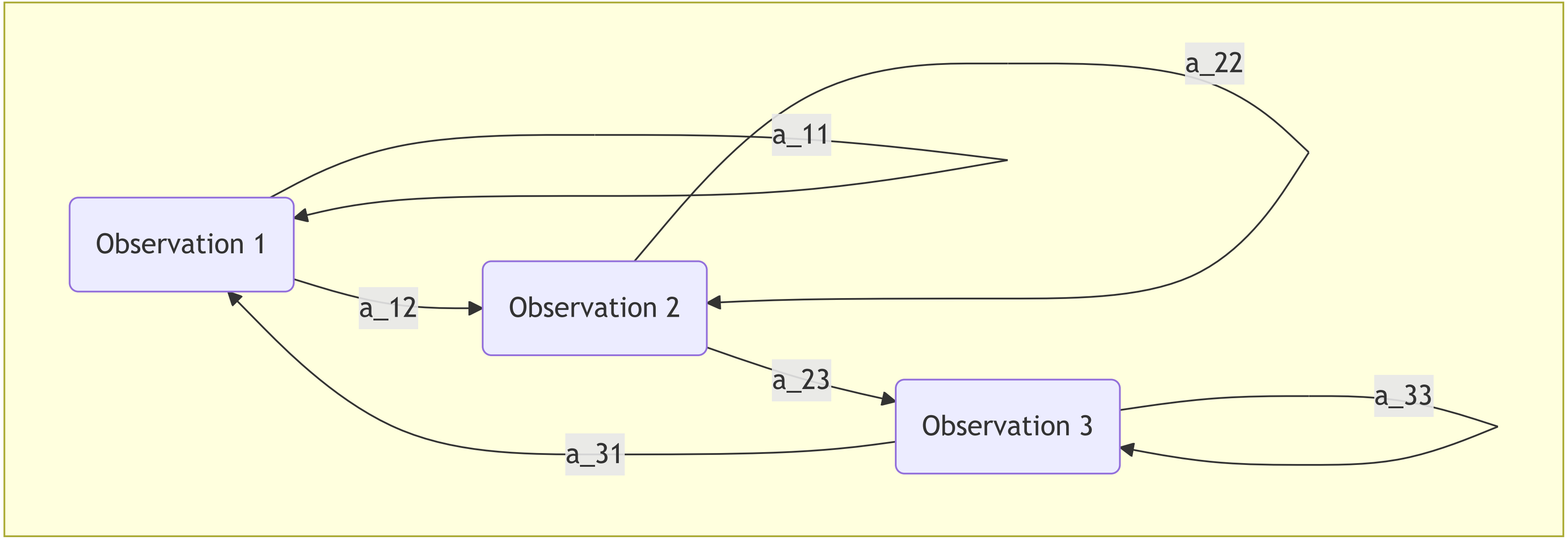
Source: The actual phonemes or spoken words being uttered by a speaker

Observed: while the acoustic signal features or audio measurements recorded by a microphone.

- The observed signal may have noise so we need to build a signal model for the source
  - to design a system to remove noise and any transmission distortion
  - to learn about the signal source (process that produced the signal) via simulation

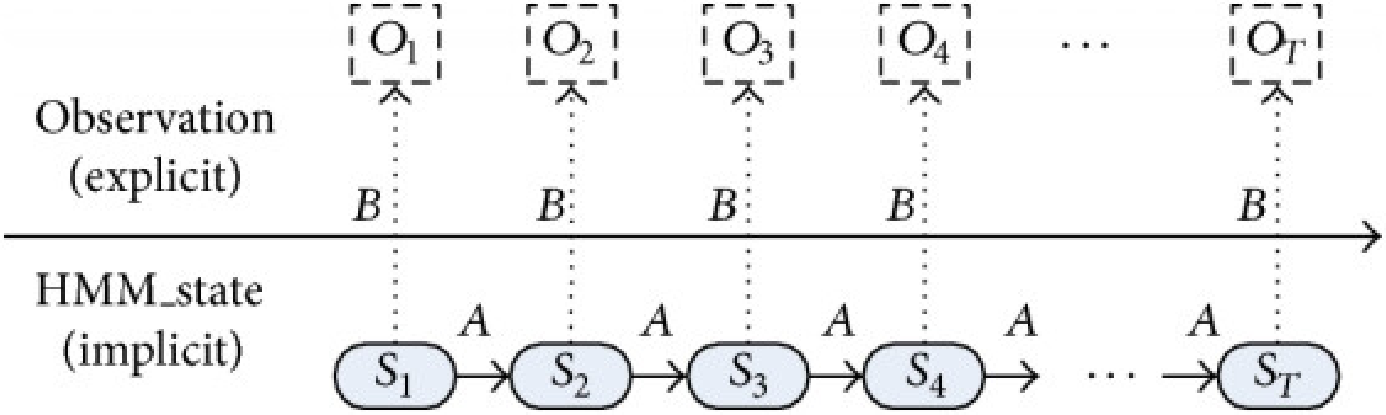
(Observable) Markov Models

- model is the set of states and transition probability matrix
- we can compute $p(O|model)$, where O is the sequence of observation

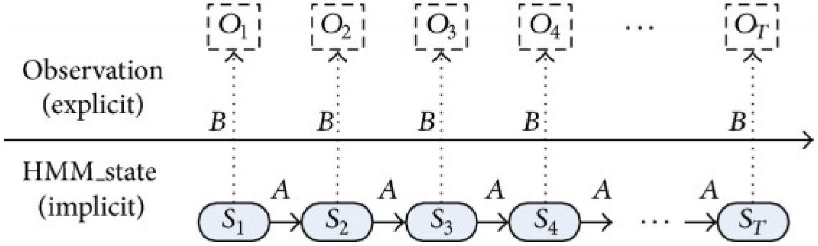


Hidden Markov Models (HMM)

- Double stochastic processes (one is hidden which is responsible for generating the process which is observable)



Elements of HMM



- states (hidden) S
- alphabets (observation symbols) O
- transition probability distribution A
- observation symbol probability (emission probability) B
- initial state distribution π

HMM Fundamental Problems: Speech Recognition Example

HMM Fundamental Problems: Speech Recognition Example

1. **Evaluation Problem**
   - Given an HMM model for a word (states = phonemes, probabilities learned)
   - Given an observed acoustic sequence o (MFCC features)
   - Compute likelihood $p(o \mid \text{model})$ that the word model generated the audio

Evaluate models on audio input to score likelihood

HMM Fundamental Problems: Speech Recognition Example

1. **Evaluation Problem**
   - Given an HMM model for a word (states = phonemes, probabilities learned)
   - Given an observed acoustic sequence o (MFCC features)
   - Compute likelihood $p(o \mid \mathrm{model})$ that the word model generated the audio

Evaluate models on audio input to score likelihood

3. **Learning Problem**
   - Given training acoustic sequences without phoneme labels
   - Estimate transition and emission probabilities to maximize data likelihood (Baum-Welch)

Learn model parameters from training data to improve accuracy

HMM Fundamental Problems: Speech Recognition Example

1. **Evaluation Problem**
   - Given an HMM model for a word (states = phonemes, probabilities learned)
   - Given an observed acoustic sequence o (MFCC features)
   - Compute likelihood $p(o \mid \text{model})$ that the word model generated the audio

Evaluate models on audio input to score likelihood

2. **Decoding Problem**
   - Given model and observed audio features
   - Find most likely phoneme state sequence that produced the audio (Viterbi)

Decode best phoneme path to recognize speech content

3. **Learning Problem**
   - Given training acoustic sequences without phoneme labels
   - Estimate transition and emission probabilities to maximize data likelihood (Baum-Welch)

Learn model parameters from training data to improve accuracy

Fundamental Problems for HMM

1. **Evaluation problem**: Compute $p(o|\lambda)$ the probability of the observation sequence given the model $\lambda = \{A, B, \pi\}$

2. **Decoding problem**: Estimate the state sequence given observation sequence and the model

3. **Learning problem**: Learn model parameters, i.e. $\arg\max_\lambda p(o|\lambda)$

Evaluation Problem of HMM

Evaluation Problem of HMM - Definition

Input

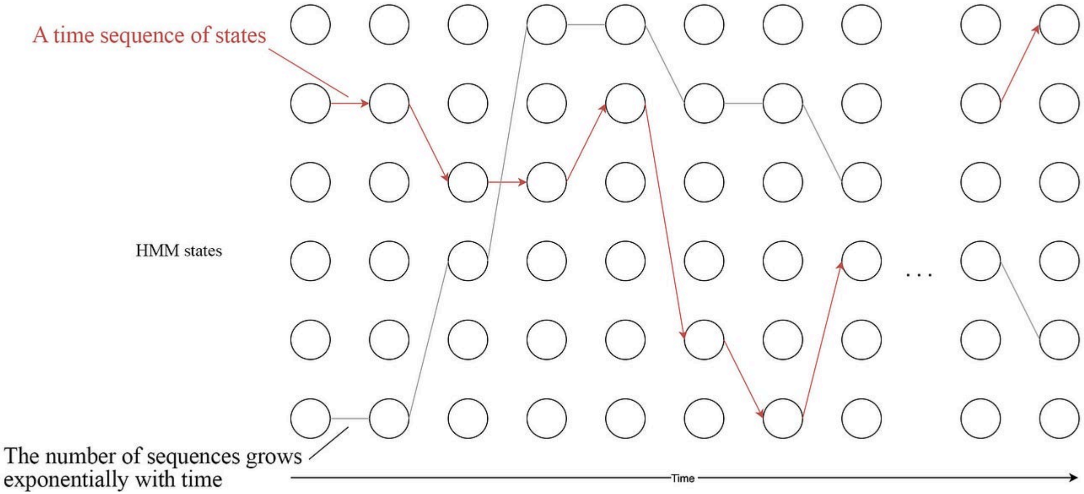
- Observation sequence $O = (o_1, o_2, \ldots, o_T)$
- HMM model $\lambda = (A, B, \pi)$ where
  - $A = \{a_{ij}\}$ transition probabilities
  - $B = \{b_j(o_t) = p(o_t|s_j)\}$ emission probabilities
  - $\pi = \{\pi_i\}$ initial state distribution

Output

The likelihood $P(O \mid \lambda)$ which is the probability that model λ produces O

Evaluation Problem of HMM - Direct Computation

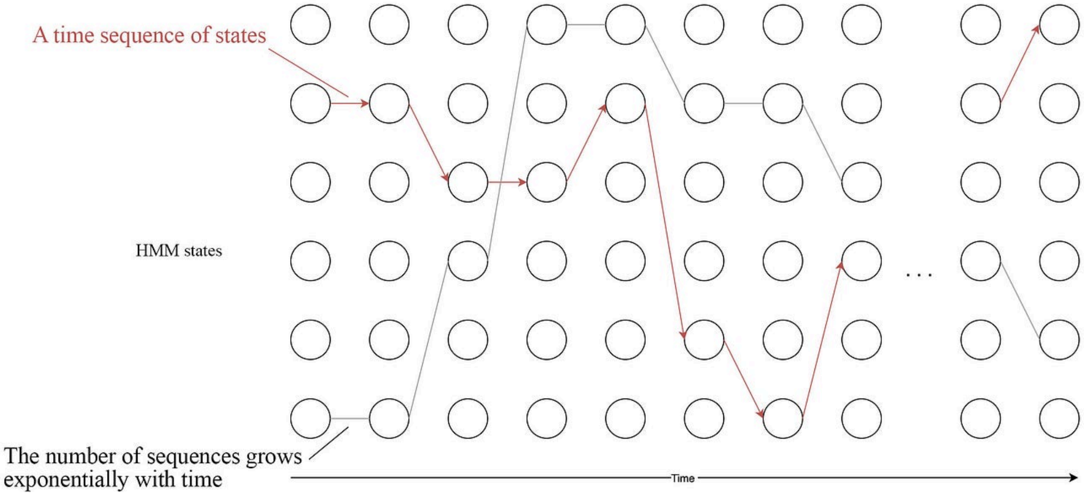
- The observation could be generated from any possible state sequence $Q = (q_1, q_2, \ldots, q_T)$



There are N^T possible state sequences

Evaluation Problem of HMM - Direct Computation

- The observation could be generated from any possible state sequence $Q = (q_1, q_2, \ldots, q_T)$



There are N^T possible state sequences

Evaluation Problem of HMM - Direct Computation

Evaluation Problem of HMM - Direct Computation

$$P(O \mid Q, \lambda) = b_1(o_1)\ .\ b_2(o_2)\ .\ b_3(o_3) \dots b_T(o_T)$$

$$P(Q \mid \lambda) = \pi_1\ .\ a_{12}\ .\ a_{23}\ .\ a_{34} \dots a_{(T-1)T}$$

Then $P(O, Q \mid \lambda) = \pi_1 b_1(o_1) \prod_{t=2}^{T} a_{(t-1)t} b_t(o_t)$

Evaluation Problem of HMM - Direct Computation

$$P(O \mid Q, \lambda) = b_1(o_1)\ .\ b_2(o_2)\ .\ b_3(o_3) \dots b_T(o_T)$$

$$P(Q \mid \lambda) = \pi_1\ .\ a_{12}\ .\ a_{23}\ .\ a_{34} \dots a_{(T-1)T}$$

Then $P(O, Q \mid \lambda) = \pi_1 b_1(o_1) \prod_{t=2}^{T} a_{(t-1)t} b_t(o_t)$

- Initially we are at time 1, we are in state q_1 with probability π_1
- Generate the symbol o_1 with probability $b_1(o_1)$
- Make a transition from state q_1 to state q_2 with probability a_{12}
- Generate the symbol o_2 with probability $b_2(o_2)$
- Continue until the last state

Evaluation Problem of HMM - Direct Computation

$P(O \mid Q, \lambda) = b_1(o_1)\ .\ b_2(o_2)\ .\ b_3(o_3) \ldots b_T(o_T)$

$P(Q \mid \lambda) = \pi_1\ .\ a_{12}\ .\ a_{23}\ .\ a_{34} \ldots a_{(T-1)T}$

Then $P(O, Q \mid \lambda) = \pi_1 b_1(o_1) \prod_{t=2}^{T} a_{(t-1)t} b_t(o_t)$

- Initially we are at time 1, we are in state q_1 with probability π_1
- Generate the symbol o_1 with probability $b_1(o_1)$
- Make a transition from state q_1 to state q_2 with probability a_{12}
- Generate the symbol o_2 with probability $b_2(o_2)$
- Continue until the last state

There are N^T possible state sequences, each requires $2T$ calculations → exponential complexity

Can we find a faster algorithm to compute the likelihood?

Learn More

Course Homepage