Hidden Markov Models

A Summary for

"A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, by Lawrence R. Rabiner"

By Seçil Öztürk

Outline

- Signal Models
- Markov Chains
- Hidden Markov Models
- Fundamentals for HMM Design
- Types of HMMs

Signal Models...

- Are used to characterize real world signals.
- Provide a basis for a theoretical description of a signal processing system.
- Tell about the signal source without having the source available.
- Are used to realize practical systems efficiently.

2 Types of Signal Models:

- Deterministic Models:
  - Specific properties of the signal are known.
    - eg. The signal is a sine wave
  - Determining values for parameters of the signal, such as frequency, amplitude, etc is required.

- Statistical Models:
  - eg. Gaussian processes, Markov processes, Hidden Markov processes
  - Characterizing the statistical properties of the signal is required.
  - Assumption:
    * Signal can be characterized as a parametric random process.
    * Parameters of the random process can be determined in a precise and well defined manner.

Discrete Markov Processes

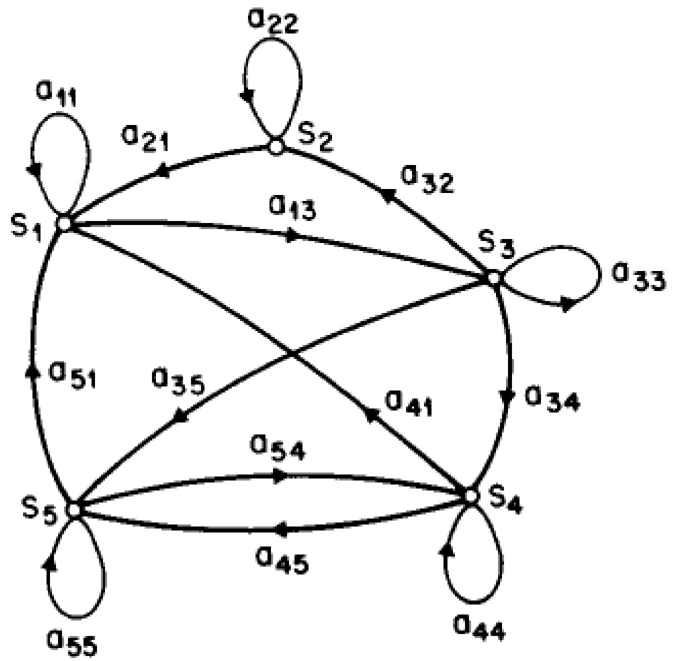


Fig. 1. A Markov chain with 5 states (labeled S_1 to S_5) with selected state transitions. [1]

- The system is described by N distinct states: $S_1, S_2 \ldots S_N$
- The system can be in one of these states at any time.
- Time instants associated with state changes are: t = 1, 2, …
- Actual state at time t is q_t
- Predecessor states must also be known for the probabilistic description.
- a_{ij}'s are state transition probabilities.

Assuming discrete, first order Markov Chain, the probabilistic description of this system is:

$$P[q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots] = P[q_t = S_j \mid q_{t-1} = S_i]$$

A 3 State Example for Weather

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

- States are defined as:
  - State 1: rainy/snowy
  - State 2: cloudy
  - State 3: sunny
- The weather at day t should be in one the states above.
- State transition matrix is A.
- a_{ij}'s represent the probabilities of going from state i to j.
- The observation sequence is denoted with O.
  - Say for t=1, sun is observed. (initial state)
  - Next observation: sun-sun-rain-rain-cloudy-sun
  - $O = \{S_3, S_3, S_1, S_1, S_3, S_2, S_3\}$
    corresponding to
    t=1,2,3,4,5,6,7,8

A 3 State Example for Weather

- The probability of the observation sequence given the model is as follows:

$$
\begin{aligned}
P(O|\text{Model}) &= P[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3|\text{Model}] \\
&= P[S_3] \cdot P[S_3|S_3] \cdot P[S_3|S_3] \cdot P[S_1|S_3] \\
&\quad \cdot P[S_1|S_1] \cdot P[S_3|S_1] \cdot P[S_2|S_3] \cdot P[S_3|S_2] \\
&= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
&= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$

where we use the notation

$$\pi_i = P[q_1 = S_i], \qquad 1 \le i \le N$$

here, π_i's are the initial state probabilities.

Hidden Markov Models

* In Markov Models,
states corresponded to observable/pyhsical events.

* In Hidden Markov Models,
observations are probabilistic functions of the state.

- So, HMMs are **doubly embedded stochastic processes**.
- The underlying stochastic process is not observable/hidden.
  It can be observed through another set of stochastic processes
  producing the observation sequences.

(ie., in Markov Models, the problem is finding the probability of the observation to be in a certain state,
in HMMs, the problem is still finding the probability of the observation to be in a certain state, but observation is also a probabilistic function of the state.)

eg. Hidden Coin Tossing Experiment, Urn and Ball Model

Elements of an HMM

- N: # of states
  Individual states: $S = \{S_1, \dots, S_N\}$ State at time t: q_t
- M: (# of distinct observation symbols)/state
  ie. discrete alphabet size in speech processing
  eg. heads & tails in coins experiment
  individual symbols: $V = \{v_1, v_2, \dots, v_M\}$
- A: State transition prob. distribution
  $aij = P[q_{t+1} = S_j \mid q_t = S_i]$ where $1 <= i, j <= N$
- B: the observation symbol probability distribution in state j
  $B = b_j(k)$ where $b_j(k) = P[v_k \text{ at } t \mid q_t = S_j]$ where $1 \le j \le N$, $1 \le k \le M.$
  eg. The probability of heads of a certain coin at time t.
- $\pi = \{ \pi_i \}$ is are the initial state distribution.
- $\pi_i = P[q_1 = S_i]$ where $1 \le i \le N$

*** Given N, M, A, B, π, HMM can be generated for O.

- O : Observation sequence $O = O_1, O_2, \dots, O_T$. O_T's are one of v_i's. T is # of total observations.

Complete specification of an HMM requires:

* A, B, π: probability measures
* N and M: model parameters
* O: observation symbols

HMM notation: $\lambda\ (A, B, \pi)$

Three Fundamental Questions in Modelling HMMs

1) Evaluation Problem:

Given Observation sequence: $O = O_1\ O_2\ \ldots\ O_T$

HMM model: $\lambda\ (A,B,\pi)$

How to compute $P(O|\lambda)$?

2) Uncover Problem:

Given Observation sequence: $O = O_1\ O_2\ \ldots\ O_T$

HMM model: $\lambda\ (A,B,\pi)$

How to choose corresponding optimal state seq. $Q=q_1 q_2 \ldots q_T$?

3) Training Problem:

How to adjust parameters A,B,π to maximize $P(O|\lambda)$?

Solution for Problem 1

$P(O|\lambda)=?$

Enumerate every possible T length state sequence.

eg. Assume fixed Q=q1q2..........qT

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda)$$

$$P(O|Q,\lambda) = b_{q_1}(O_1)\cdot b_{q_2}(O_2)\cdots b_{q_T}(O_T).$$

$$P(Q|\lambda) = \pi_{q_1}a_{q_1q_2}a_{q_2q_3}\cdots a_{q_{T-1}q_T}.$$

$$\begin{aligned} P(O|\lambda) &= \sum_{\text{all } Q} P(O|Q,\lambda)\, P(Q|\lambda) \\ &= \sum_{q_1,q_2,\ldots,q_T} \pi_{q_1}b_{q_1}(O_1)\, a_{q_1q_2}b_{q_2}(O_2) \\ &\quad \cdots a_{q_{T-1}q_T}b_{q_T}(Q_T). \end{aligned}$$

Unfeasible computation time!

On order of $2T.N^T$

Solution for Problem 1

Forward-Backward procedure

Forward variable:

$\alpha_t(i)=P(O_1O_2....O_t, q_t=S_i|\lambda)$

(prob. for partial observation sequence $O_1...O_t$ ending at state S_i at time t, λ)

Inductive Solution!

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \leq i \leq N.$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \leq t \leq T-1$$

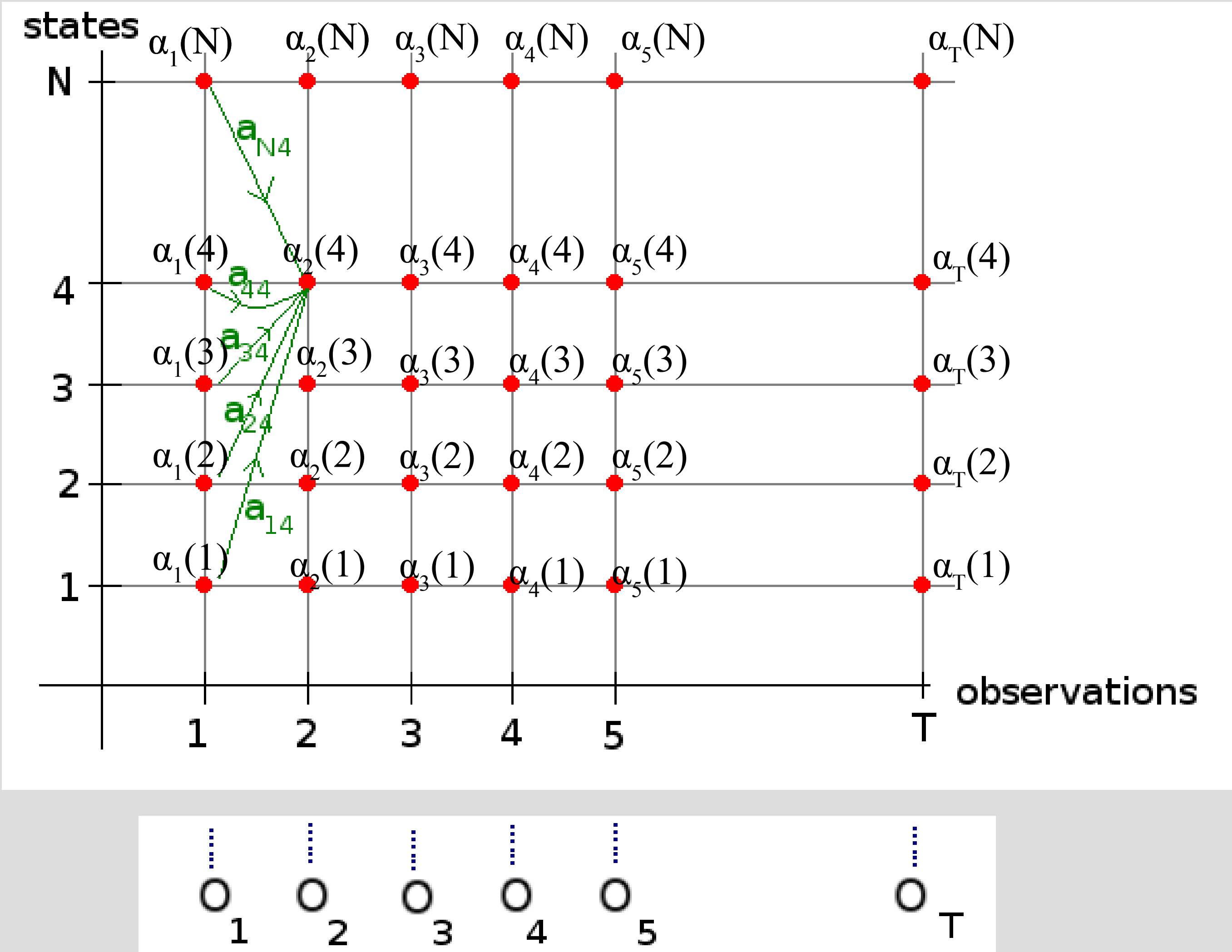
$$1 \leq j \leq N.$$

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i).$$

Computation time:

On order of N^2T

Trellis


states
observations
N
4
3
2
1
1
2
3
4
5
T
α1(N) α2(N) α3(N) α4(N) α5(N) αT(N)
α1(4) α2(4) α3(4) α4(4) α5(4) αT(4)
α1(3) α2(3) α3(3) α4(3) α5(3) αT(3)
α1(2) α2(2) α3(2) α4(2) α5(2) αT(2)
α1(1) α2(1) α3(1) α4(1) α5(1) αT(1)
aN4
a44
a34
a24
a14
O1 O2 O3 O4 O5 OT

Backward Variable:

$\beta_t(i)=P(O_{t+1}O_{t+2}\ldots O_T|q_T=S_i,\lambda)$

(probability of the partial observation sequence from t+1 to end, given state Si at time t, λ)

Inductive Solution!

$$\beta_T(i) = 1, \qquad 1 \le i \le N.$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1})\, \beta_{t+1}(j),$$
$$t = T-1, T-2, \cdots, 1, \; 1 \le i \le N.$$

Computation time:

On order of N^2T

Solution for Problem 2

* Aim is to find an optimal state sequence for the observation.
* Several solutions exist.
* Optimality criteria must be adjusted.

eg. states individually most likely at time t.
maximizes expected # of correct individual states.

A posteriori probability variable γ:

$\gamma_t(i)=P(q_t=S_i|O,\lambda)$

(probability of being in state S_i at time t given observation sequence O and λ)

$$\gamma_t(i) = P(q_t = i \mid \mathbf{O}, \lambda) = \frac{P(\mathbf{O}, q_t = i \mid \lambda)}{P(\mathbf{O} \mid \lambda)} = \frac{P(\mathbf{O}, q_t = i \mid \lambda)}{\sum_{i=1}^{N} P(\mathbf{O}, q_t = i \mid \lambda)}.$$

$$\gamma_t(i) = \frac{\alpha_t(i)\,\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\,\beta_t(i)}$$

$P(O|\lambda)$ is normalization factor to make sure sum of $\gamma_t(i)$'s equals 1

Individually most likely state q_t at time t:

$$q_t = \operatorname*{argmax}_{1 \le i \le N} [\gamma_t(i)], \qquad 1 \le t \le T.$$

Problem:

This equation finds the most likely state at each t regardless of the probability of occurrence of states, so the resulting sequence may be invalid.

Possible solution to the problem above:
Find the state sequence maximizing pairs or triples of states

OR

Find the single best state sequence
to maximize $P(Q|O,\lambda)$
equivalent to maximize $P(Q,O|\lambda)$

Viterbi Algorithm

Aim: to find the single best state sequence $Q=\{q_1 q_2 \ldots q_T\}$
for given observation sequence $O=\{O_1 O_2 \ldots O_T\}$

Define δ:

$$\delta_t(i) = \max_{q_1, q_2, \cdots, q_{t-1}} P[q_1\, q_2 \cdots q_t = i,\, O_1\, O_2 \cdots O_t | \lambda]$$

(the best score, ie. highest probability along a single path, at time t)
(accounts for the first t observations, ends in state Si)

$$\delta_{t+1}(j) = [\max_i \delta_t(i)\, a_{ij}] \cdot b_j(O_{t+1}).$$

For each t and j, must keep track of argument maximizing above equation.

Use array $\psi_t(j)$

Viterbi Algorithm

To find the best state sequence:

1. Initialization

$$\delta_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N$$
$$\psi_1(i) = 0.$$

2. Recursion

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \qquad 2 \le t \le T, \quad 1 \le j \le N$$
$$\psi_t(j) = \operatorname*{argmax}_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \qquad 2 \le t \le T, \quad 1 \le j \le N.$$

3. Termination

$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$
$$q_T^* = \operatorname*{argmax}_{1 \le i \le N} [\delta_T(i)].$$

4. Path/State Sequence Backtracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, \cdots, 1.$$

- Just like forward procedure.
- But finds max instead of summation.
- ψ Keeps track of maximizing points

Solution for Problem 3

Aim: Adjusting A, B, π to maximize the probability of the training data.

Choose λ (A,B,π) such that $P(O|\lambda)$ is locally maximized using:

Methods:
* Baum-Welch Method
* Expectation-Modification (EM) Method
* Gradient Techniques

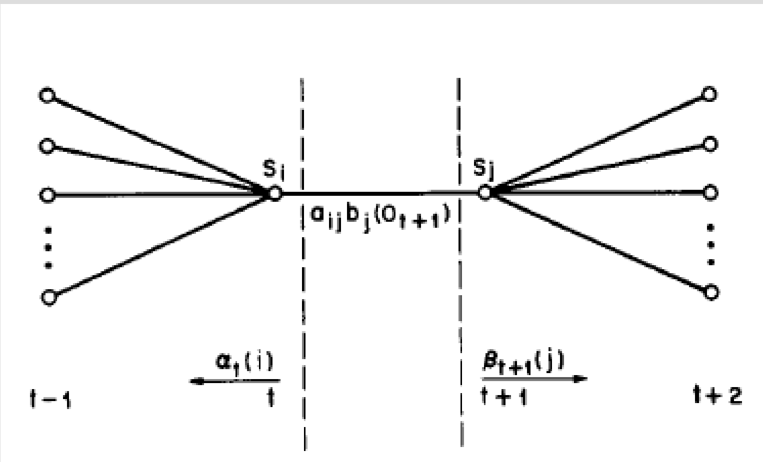
Baum-Welch Method

Define Variable ξ:

$\xi_t(i,j) = P\,(q_t = S_i, q_{t+1} = S_j | O, \lambda)$

(the probability of being in state Si at t, in Sj at t+1, given observation and model)

The path satisfying this condition:



$$\xi_t(i,j) = \frac{P(q_t = i,\, q_{t+1} = j,\, \mathbf{O} \mid \lambda)}{P(\mathbf{O} \mid \lambda)}$$

$$\xi_t(i,j) = \frac{\alpha_t(i)\, a_{ij} b_j(O_{t+1})\, \beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i)\, a_{ij} b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i)\, a_{ij} b_j(O_{t+1})\, \beta_{t+1}(j)}$$

Relate to γ:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j).$$

Baum-Welch Method

Expected number of transitions made from state Si in O: $\sum_{t=1}^{T-1} \gamma_t(i)$

Expected number of transitions made from state Si to Sj in O: $\sum_{t=1}^{T-1} \xi_t(i,j)$

Reestimation formulas for A,B,π :

$$\bar{\pi}_i = \text{expected frequency (number of times) in state } S_i \text{ at time } (t = 1) = \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } S_i \text{ to state } S_j}{\text{expected number of transitions from state } S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{b}_j(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j} = \frac{\sum_{\substack{t=1 \\ \text{s.t. } O_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}.$$

Baum-Welch Method

Current Model: $\lambda\,(A,B,\pi)$

Reestimation Model: $\bar{\lambda}\,(\bar{A}, \bar{B}, \bar{\pi},)$

Either;

1) λ defines critical point of the likelihood function, where $\lambda=\bar{\lambda}$

2) model $\bar{\lambda}$ is more likely than λ

in the sense $P(O|\bar{\lambda})>P(O|\lambda)$

So $\bar{\lambda}$ is the new model matching the observation sequence better.

Using $\bar{\lambda}$ as λ iteratively and repeating reestimation calculation, improvement for the probability of O being observed in model is reached.

Final result is called *a maximum likelihood estimate of the HMM*.

Baum-Welch Method

Reestimation formulas can be derived by maximizing Baum's auxiliary function over $\bar{\lambda}$:

$$Q(\lambda, \bar{\lambda}) = \sum_{Q} P(Q|O, \lambda) \log [P(O, Q|\bar{\lambda})]$$

Proved that maximizing $Q(\lambda, \bar{\lambda})$ leads to increased likelihood.

$$\max_{\bar{\lambda}} [Q(\lambda, \bar{\lambda})] \Rightarrow P(O|\bar{\lambda}) \geq P(O|\lambda).$$

Eventually likelihood function converges to a critical point.

Baum-Welch Method

Stochastic constraints are satisfied in each reestimation procedure:

$$\sum_{i=1}^{N} \bar{\pi}_i = 1$$

$$\sum_{j=1}^{N} \bar{a}_{ij} = 1, \qquad 1 \le i \le N$$

$$\sum_{k=1}^{M} \bar{b}_j(k) = 1, \qquad 1 \le j \le N$$

Also, Lagrange multipliers can be used to find $\pi, a_{ij}, b_j(k)$ parameters maximizing $P(O|\lambda)$
(Think of the parameter estimation as a constrained optimization problem for $P(O|\lambda)$, constrained by above equations)

Using Lagrange Multipliers;

$$\pi_i = \frac{\pi_i \dfrac{\partial P}{\partial \pi_i}}{\sum_{k=1}^{N} \pi_k \dfrac{\partial P}{\partial \pi_k}}$$

$$a_{ij} = \frac{a_{ij} \dfrac{\partial P}{\partial a_{ij}}}{\sum_{k=1}^{N} a_{ik} \dfrac{\partial P}{\partial a_{ik}}}$$

$$b_j(k) = \frac{b_j(k) \dfrac{\partial P}{\partial b_j(k)}}{\sum_{\ell=1}^{M} b_j(\ell) \dfrac{\partial P}{\partial b_j(\ell)}}.$$

Manipulating these equations, it can be shown that reestimation formulas are correct at critical points of $P(O|\lambda)$

Types of HMMs

So far, considered only ergodic HMMs:

<u>Ergodic Model:</u>

Every state transition is possible. a_{ij}'s positive.

<u>Left-Right (Bakis) Model:</u>

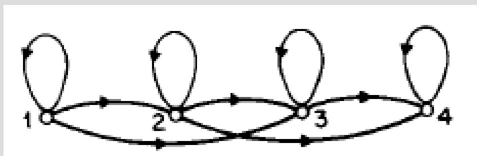
As time increases, state index increases or stays the same. ie. states proceed from left to right.

$$a_{ij} = 0, \quad j < i$$

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

$$a_{NN} = 1$$

$$a_{Ni} = 0, \quad i < N.$$



$$a_{ij} = 0, \quad j > i + \Delta$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}.$$

Continuous Observation Densities in HMMs

- Finite alphabet up to now.
- Observations are continuous signals/vectors.
- General representation of the pdf:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} \mathfrak{N}[\mathbf{O}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm}], \qquad 1 \le j \le N$$

  - O: vector being modeled
  - c_{jm}: mixture coeff. for m^{th} mixture in state j
  - $\mathfrak{N}$ log concave/elliptical symmetric density (eg. Gaussian) mean: μ_{jm}, cov: U_{jm}
- c_{jm} should satisfy

$$\sum_{m=1}^{M} c_{jm} = 1, \qquad 1 \le j \le N$$
$$c_{jm} \ge 0, \qquad 1 \le j \le N, 1 \le m \le M$$

such that pdf is normalized: $\int_{-\infty}^{\infty} b_j(x)\, dx = 1, \qquad 1 \le j \le N.$

Continuous Observation Densities in HMMs

Reestimation formulas:

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T}\sum_{k=1}^{M} \gamma_t(j,k)}$$

$$\bar{\boldsymbol{\mu}}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot \mathbf{O}_t}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

$$\bar{\mathbf{U}}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot (\mathbf{O}_t - \boldsymbol{\mu}_{jk})(\mathbf{O}_t - \boldsymbol{\mu}_{jk})'}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

$\gamma_t(j,k)$ prob. Of being in state j at time t with k^{th} mixture component accounting for O_t

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\,\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\,\beta_t(j)}\right]\left[\frac{c_{jk}\,\mathfrak{N}(\mathbf{O}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk})}{\sum_{m=1}^{M} c_{jm}\,\mathfrak{N}(\mathbf{O}_t, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm})}\right].$$

Autoregressive HMMs

- Particularly applicable to speech processing.
- Observation vectors are drawn drom an autoregression process.
- Observation vector O: $(x_0, x_1, \ldots, x_{k-1})$
- Ok's are related by: $O_k = -\sum_{i=1}^{p} a_i O_{k-i} + e_k$

  where e_k, k=0, 1, 2, 3, ..., p are Gaussian, independent, identically distributed rv. with zero mean, variance σ^2
- a_{ij}, i=1,...,p are predictor (autoregression) coefficients.

Autoregressive HMMs

- For large K, density function O is approximately:

$$f(\mathbf{O}) = (2\pi\sigma^2)^{-K/2} \exp\left\{-\frac{1}{2\sigma^2}\,\delta(\mathbf{O}, \mathbf{a})\right\}$$

where

$$\delta(\mathbf{O}, \mathbf{a}) = r_a(0)\, r(0) + 2\sum_{i=1}^{p} r_a(i)\, r(i)$$

$$\mathbf{a}' = [1, a_1, a_2, \cdots, a_p]$$

$$r_a(i) = \sum_{n=0}^{p-i} a_n a_{n+i} \qquad (a_0 = 1), 1 \le i \le p$$

$$r(i) = \sum_{n=0}^{K-i-1} x_n x_{n+i} \qquad 0 \le i \le p.$$

- r(i) autocorrelation of observation samples
- r_a(i) autocorr. Of autoreg. coeff.s

Autoregressive HMMs

- Total prediction residual α is

$$\alpha = E\left[\sum_{i=1}^{K} (e_i)^2\right] = K\sigma^2$$

σ^2 is variance per sample of error signal.

- Normalized observation vector: $$\hat{O} = \frac{O}{\sqrt{\alpha}} = \frac{O}{\sqrt{K\sigma^2}}$$

- Samples xi's are divided by $\sqrt{K\sigma^2}$ (normalized by sample variance)

- $$f(\hat{O}) = \left(\frac{2\pi}{K}\right)^{-K/2} \exp\left(-\frac{K}{2}\,\delta(\hat{O}, a)\right).$$

Autoregressive HMMs

Using Gaussian autoregressive density, assume the mixture density:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} b_{jm}(\mathbf{O})$$

Each b_{jm}(O) is denstiy with autoregression vector ajm (or autocorr.vector r_{ajm})

$$b_{jm}(\mathbf{O}) = \left(\frac{2\pi}{K}\right)^{-K/2} \exp\left\{-\frac{K}{2}\,\delta(\mathbf{O}, \mathbf{a}_{jm})\right\}.$$

Reestimation formula for sequence autocorrelation r(i) for the jth state, kth mixture component:

$$\bar{\mathbf{r}}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j, k) \cdot \mathbf{r}_t}{\sum_{t=1}^{T} \gamma_t(j, k)}$$

Where γt(j,k) is the prob. of being in state j at time t, using mixture component k,

$$\gamma_t(j, k) = \left[\frac{\alpha_t(j)\,\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\,\beta_t(j)}\right]\left[\frac{c_{jk} b_{jk}(\mathbf{O}_t)}{\sum_{k=1}^{M} c_{jk} b_{jk}(\mathbf{O}_t)}\right].$$

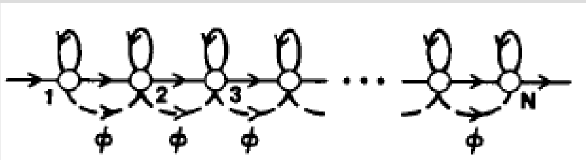
Null Transitions

NULL Transitions:

Observations are associated with the arcs of the model.

Used for transitions which makes no output. (jumps between states produce no observation)

Eg: a left-right model:



It is possible to omit transitions between states and conclude with 1 observation to account for a path beginning in state 1, ending in state N.

Tied States

- Equivalence relation between HMM parameters in different states.
- # of independent parameters in model is reduced.
- Used in cases where observation density is the same for two or more states. (eg in speech sounds)
- Model becomes simpler for parameter estimation

More...

- Inclusion of Explicit State Duration Density in HMMs
- Optimization Criterion

Bibliography

- A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, by Lawrence R. Rabiner
- Fundamentals of Speech Recognition, by Lawrence R. Rabiner Biign Hwang Juang

Thanks for Listening!