Hidden Markov Models

Tapas Kanungo

Language and Media Processing Lab

Center for Automation Research

University of Maryland

Web: www.cfar.umd.edu/~kanungo

Email: kanungo@cfar.umd.edu

March 16, 1998

Outline

1. **Markov models**
2. **Hidden Markov models**
3. **Forward/Backward algorithm**
4. **Viterbi algorithm**
5. **Baum-Welch estimation algorithm**

Markov Models

- **Observable states:**

$$1, 2, \ldots, N$$

- **Observed sequence:**

$$q_1, q_2, \ldots, q_t, \ldots, q_T$$

- **First order Markov assumption:**

$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \ldots) = P(q_t = j | q_{t-1} = i)$$

- **Stationarity:**

$$P(q_t = j | q_{t-1} = i) = P(q_{t+l} = j | q_{t+l-1} = i)$$

Markov Models

- **State transition matrix A :**

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2N} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{iN} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{Nj} & \cdots & a_{NN} \end{bmatrix}$$

where

$$a_{ij} = P(q_t = j | q_{t-1} = i) \qquad 1 \leq i, j, \leq N$$

- **Constraints on a_{ij} :**

$$\begin{aligned} a_{ij} &\geq 0, \qquad \forall i, j \\ \sum_{j=1}^{N} a_{ij} &= 1, \qquad \forall i \end{aligned}$$

Markov Models: Example

- **States:**

  1. **Rainy (R)**
  2. **Cloudy (C)**
  3. **Sunny (S)**

- **State transition probability matrix:**

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

- **Compute the probability of observing $SSRRSCS$ given that today is S.**

Markov Models: Example

Basic conditional probability rule:

$$P(A, B) = P(A|B)P(B)$$

The Markov chain rule:

$$\begin{aligned}
&P(q_1, q_2, \ldots, q_T) \\
&= P(q_T|q_1, q_2, \ldots, q_{T-1})P(q_1, q_2, \ldots, q_{T-1}) \\
&= P(q_T|q_{T-1})P(q_1, q_2, \ldots, q_{T-1}) \\
&= P(q_T|q_{T-1})P(q_{T-1}|q_{T-2})P(q_1, q_2, \ldots, q_{T-2}) \\
&= P(q_T|q_{T-1})P(q_{T-1}|q_{T-2}) \cdots P(q_2|q_1)P(q_1)
\end{aligned}$$

Markov Models: Example

- Observation sequence O :

$$O = (S, S, S, R, R, S, C, S)$$

- Using the chain rule we get:

$$\begin{aligned}
&P(O|model) \\
&= P(S, S, S, R, R, S, C, S|model) \\
&= P(S)P(S|S)P(S|S)P(R|S)P(R|R) \times \\
&\quad P(S|R)P(C|S)P(S|C) \\
&= \pi_3 a_{33} a_{33} a_{31} a_{11} a_{13} a_{32} a_{23} \\
&= (1)(0.8)^2(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}$$

- The prior probability $\pi_i = P(q_1 = i)$

Markov Models: Example

- **What is the probability that the sequence remains in state i for exactly d time units?**

$$\begin{aligned} p_i(d) &= P(q_1 = i, q_2 = i, \ldots, q_d = i, q_{d+1} \neq i, \ldots) \\ &= \pi_i (a_{ii})^{d-1}(1 - a_{ii}) \end{aligned}$$

- **Duration density of Markov chains is exponential.**
- **What is the expected value of the duration d in state i?**

$$\begin{aligned} \bar{d}_i &= \sum_{d=1}^{\infty} d p_i(d) \\ &= \sum_{d=1}^{\infty} d (a_{ii})^{d-1}(1 - a_{ii}) \\ &= (1 - a_{ii}) \sum_{d=1}^{\infty} d (a_{ii})^{d-1} \\ &= (1 - a_{ii}) \frac{\partial}{\partial a_{ii}} \sum_{d=1}^{\infty} (a_{ii})^d \\ &= (1 - a_{ii}) \frac{\partial}{\partial a_{ii}} \left(\frac{a_{ii}}{1 - a_{ii}} \right) \\ &= \frac{1}{1 - a_{ii}} \end{aligned}$$

Markov Models: Example

- Avg. number of consecutive sunny days =

$$\frac{1}{1 - a_{33}} = \frac{1}{1 - 0.8} = 5$$

- Avg. number of consecutive cloudy days = 2.5

- Avg. number of consecutive rainy days = 1.67

Hidden Markov Models

- **States are not observable**
- **Observations are probabilistic functions of state**
- **State transitions are still probabilistic**

Urn and Ball Model

- N urns containing colored balls
- M distinct colors of balls
- Each urn has a (possibly) different distribution of colors
- Sequence generation algorithm:
  1. Pick initial urn according to some random process.
  2. Randomly pick a ball from the urn and then replace it
  3. Select another urn according a random selection process associated with the urn
  4. Repeat steps 2 and 3

Elements of Hidden Markov Models

- N – the number of hidden states
- Q – set of states $Q = \{1, 2, \ldots, N\}$
- M – the number of symbols
- V – set of symbols $V = \{1, 2, \ldots, M\}$
- A – the state-transition probability matrix.

$$a_{ij} = P(q_{t+1} = j | q_t = i) \quad 1 \leq i, j, \leq N$$

- B – Observation probability distribution:

$$B_j(k) = P(o_t = k | q_t = j) \quad 1 \leq k \leq M$$

- π – the initial state distribution:

$$\pi_i = P(q_1 = i) \quad 1 \leq i \leq N$$

- λ – the entire model $\lambda = (A, B, \pi)$

Three Basic Problems

1. Given observation $O = (o_1, o_2, \ldots, o_T)$ and model $\lambda = (A, B, \pi)$, efficiently compute $P(O|\lambda)$.
   - Hidden states complicate the evaluation
   - Given two models λ_1 and λ_2, this can be used to choose the better one.
2. Given observation $O = (o_1, o_2, \ldots, o_T)$ and model λ find the optimal state sequence $q = (q_1, q_2, \ldots, q_T)$.
   - Optimality criterion has to be decided (e.g. maximum likelihood)
   - "Explanation" for the data.
3. Given $O = (o_1, o_2, \ldots, o_T)$, estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$.

Solution to Problem 1

- Problem: Compute $P(o_1, o_2, \ldots, o_T|\lambda)$
- Algorithm:
  - Let $q = (q_1, q_2, \ldots, q_T)$ be a state sequence.
  - Assume the observations are independent:

$$\begin{aligned} P(O|q,\lambda) &= \prod_{i=1}^{T} P(o_t|q_t,\lambda) \\ &= b_{q_1}(o_1) b_{q_2}(o_2) \cdots b_{q_T}(o_T) \end{aligned}$$

  - Probability of a particular state sequence is:

$$P(q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}$$

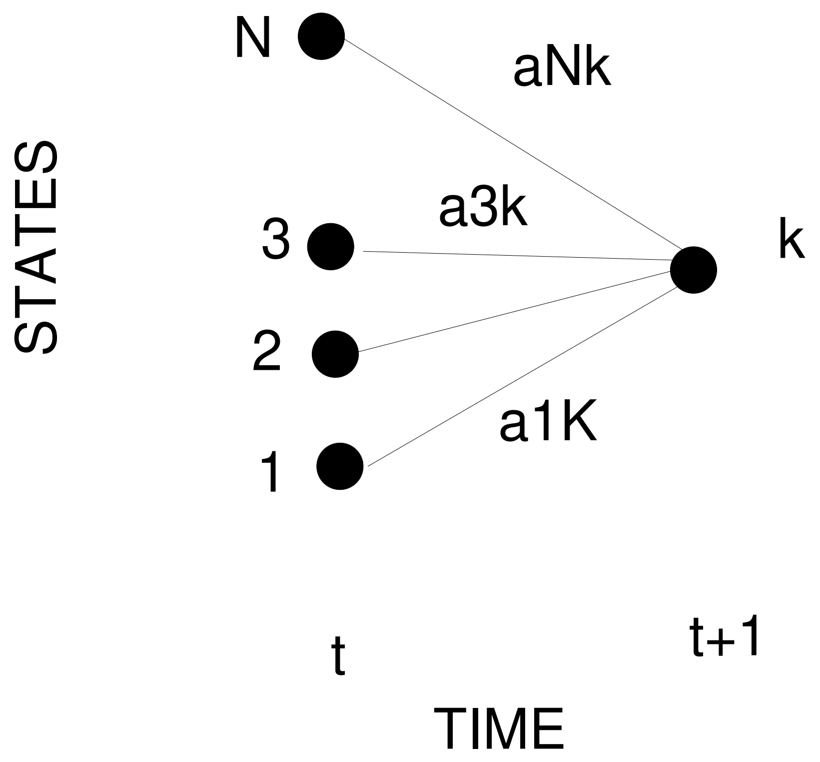
  - Also, $P(O, q|\lambda) = P(O|q,\lambda) P(q|\lambda)$
  - Enumerate paths and sum probabilities:

$$P(O|\lambda) = \sum_q P(O|q,\lambda) P(q|\lambda)$$

- N^T state sequences and $O(T)$ calculations.

  Complexity: $O(TN^T)$ calculations.

Forward Procedure: Intuition



Forward Procedure

- Define forward variable $\alpha_t(i)$ as:

$$\alpha_t(i) = P(o_1, o_2, \ldots, o_t, q_t = i|\lambda)$$

- $\alpha_t(i)$ is the probability of observing the partial sequence $(o_1, o_2 \ldots, o_t)$ such that the state q_t is i.
- Induction:

  1. Initialization: $\alpha_1(i) = \pi_i b_i(o_1)$
  2. Induction:

  $$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1})$$

  3. Termination:

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

- Complexity: $O(N^2T)$.

Homework: Understanding HMM

Consider the following coin-tossing experiment:

	State 1	State 2	State 3
P(H)	0.5	0.75	0.25
P(T)	0.5	0.25	0.25

– state-transition probabilities equal to 1/3

– initial state probabilities equal to 1/3

1. You observe $O = (H, H, H, H, T, H, T, T, T, T)$. What state sequence, q, is most likely? What is the joint probability, $P(O, q|\lambda)$, of the observation sequence and the state sequence?
2. What is the probability that the observation sequence came entirely of state 1?

3. **Consider the observation sequence**

$$\tilde{O} = (H, T, T, H, T, H, H, T, T, H).$$

How would your answers to parts 1 and 2 change?

4. **If the state transition probabilities were:**

$$A' = \begin{bmatrix} 0.9 & 0.45 & 0.45 \\ 0.05 & 0.1 & 0.45 \\ 0.05 & 0.45 & 0.1 \end{bmatrix},$$

how would the new model λ' change your answers to parts 1-3?

5. **Note: Do not use the Viterbi algorithm for the above problems.**